The Laguerre mixture model.

Radcliffe, John Edward

The Laguerre Mixture Model

by

John Edward Radcliffe

Thesis

submitted to the University of Wales in candidature for the degree of

MASTER OF PHILOSOPHÆ

European Business Management School
University of Wales, Swansea
Swansea SA2 8PP
United Kingdom

February 2005

Declaration

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

John Edward Radcliffe
February 2005

Statement

This thesis is the result of my own investigation, except where acknowledgement of other sources is given.

John Edward Radcliffe
February 2005

Statement

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan and for the title and summary to be made available to outside organisations. Copyright is invested in the author of the thesis.

John Edward Radcliffe
February 2005

Summary

In recent years, time series analysis has become a highly developed subject. However it still remains that relatively few published studies exist on non-linear modelling. This is possibly due to their complexity and vast diversity compared with linear models and therefore inability to generalise such models. In this thesis, we address this deficiency and examine the Laguerre mixture model. In particular, we examine the likelihood function of the Laguerre mixture model from a time series with missing values. Further through the medium of simulation we investigate the statistical properties of the maximum likelihood estimators for varied sample sizes and true values of the parameter of the model.

It must be noted that there has been a lapse of ten years between completion of this study and submission. Since completion there have been relevant advances in non-linear time series that are not been discussed here, for example see Rothman (1999), Franses and Dijk (2000) and Kantz and Schreiber (2003).

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor, Dr. John Pemberton, for his continued help, earnest supervision and constant encouragement throughout the duration of this dissertation. I would also like to thank Dr. Assad Jalali who in conjunction with Dr. John Pemberton devised the initial idea that inspired this thesis, and for his enlightening suggestions and constructive advice.

May I also take this opportunity to thank my family for their understanding and support. Finally, I would like to the European Business Management School postgraduate members, both past and present, and the friends that I have made for such an enjoyable and memorable time in Swansea University.

JOHN EDWARD RADCLIFFE

University of Wales, Swansea
February 2005

Table of Contents

List of Figures

List of Tables

Chapter One

Introduction

Although there exists a burgeoning empirical literature dedicated to the theory and practice of time series analysis, the majority of these studies focus on linear time series models. However with most aspects of the real world being concerned with non-linear processes, this has led to many convincing arguments for the requirement of non-linear time series analysis. Consequently in recent years we have witnessed the development of non-linear time series models. Although these models require deeper aspects of probability theory and more sophisticated tools of statistical inference, they provide the analyst with a more real approximation to the world around us.

In this thesis, we examine a new class of non-linear models known as mixture models, as developed by Jalali and Pemberton (1995). In particular, we focus on the Laguerre mixture model. We wish to examine the likelihood function of the Lageurre mixture model for a time series with missing values. Further, using box-plots, histograms and Q-Q plots, we investigate the statistical properties of the maximum likelihood estimators $\hat{r}$ and $\hat{\lambda}$, for the Laguerre mixture model from a time series with missing values.

In chapter two, we provide a brief introduction and some basic definitions of univariate time series analysis. In addition, we discuss linear modelling and the theory of maximum likelihood estimation and the case of parameter estimation for time series with missing values.

In chapter three, we discuss the development of non-linear modelling and provide several examples of the more commonly known non-linear model. Finally, we discuss the problem of parameter estimation for non-linear time series analysis.

In chapter four, we discuss the underlying theory of mixture models as developed by Jalali and Pemberton (1995). We discuss their properties and further consider the zero-order threshold model as a special case of mixture model.

In chapter five, we introduce the Laguerre mixture model. In particular, we investigate the likelihood function for a time series with missing values. In anticipation of our results we find that due to the complexity of the first derivatives of $L(\theta)$, we are unable to find a critical point analytically. Therefore we investigate whether a maximum point exists through examining plots and numerical optimisation routines. Our results indicate that a unique maximum point does exist.

In chapter six, we examine the statistical properties of the maximum likelihood estimators $\hat{r}$ and $\hat{\lambda}$, for the Laguerre mixture model from a time series with missing values. In particular we examine box-plots, histograms and Q-Q plots of both $\hat{r}$ and $\hat{\lambda}$ for specified values of r_0 and λ_0 with time series of varied lengths. We find that as the length of the time series increases the distributions of both $\hat{r}$ and $\hat{\lambda}$ approach the

normal distribution, although this does appear to be quicker for the estimators $\hat{\lambda}$ as opposed to $\hat{r}$.

The findings of this thesis are summarised in chapter seven, where conclusions are then drawn on the work as a whole.

Chapter Two

Univariate Time Series

In this chapter, we provide some elementary concepts and definitions of univariate time series. Further, we discuss the development of these models, the method of maximum likelihood estimation and parameter estimation for a time series with missing values.

2.1 Introduction

A time series is a set of observations measured at different points of time. Examples occur in a variety of fields such as records of voltages in an electrical circuit measured

at one-second intervals, daily air temperatures, the monthly price index of a commodity, or an EEG record measuring the electrical activity between two locations in a human brain. The first three of these are examples of time series recorded at a discrete set of time points i.e. every second, every day, every month, whereas the fourth is an example of a time series recorded continuously over a time period. Although both types of series are of interest, in practice we shall consider only the discrete type.

An important property that identifies time series data from other types of statistical data is the fact that observations of the series at different time instants will be dependent. For example this is demonstrated in records of daily air temperatures where one would expect today's temperature to be dependent on temperatures of previous days. When analysing a time series, a fundamental consideration is studying the pattern of the correlation between observations at different time instants, and trying to construct a statistical model that explains the correlation structure of the series.

Although many well established methods for the statistical modelling of time series data now exist. The methods developed up to the end of the first half of this century could be said to have led to the seminal work of Box and Jenkins in the sixties, culminating in their famous book in 1970. Other works emanating from this include the books by Hannan (1970), Anderson (1971), Brillinger (1975) and Priestley (1981). However, virtually all of these recognised methods for model fitting rest on two basic assumptions. Stationarity is the first assumption whilst the second is that the series corresponds to a linear model. This implies that the value of the observed series can be represented as linear combinations of present and past values of a strictly random series. The following sections and notation closely follow Priestley (1981).

2.2 Assumptions

2.2.1 Stationarity

A series $\{Z_t\}$, t=0,±1,±2,..., is said to be stationary if the statistical properties remain constant over time. More accurately speaking, $\{Z_t\}$ is said to be completely or strictly stationary if, for any set of times $t_1, t_2, ..., t_n$ and any integer k, the joint probability distribution of $\{Z_{t_1}, Z_{t_2}, ..., Z_{t_n}\}$ is the same as the joint probability distribution of $\{Z_{t_1+k}, Z_{t_2+k}, ..., Z_{t_n+k}\}$. However, this is quite a severe requirement, which is relaxed by introducing a weaker condition known as "stationarity up to order m".

A series $\{Z_t\}$ is said to be stationary up to order m if, for any set of times $t_1, t_2, ..., t_n$ and any integer k, all the joint moments up to order m of $\{Z_{t_1}, Z_{t_2}, ..., Z_{t_n}\}$ exist and equal the corresponding joint moments up to order m of $\{Z_{t_1+k}, Z_{t_2+k}, ..., Z_{t_n+k}\}$.

Thus,

$$E\left[\{Z_{t_1}\}^{m_1}\{Z_{t_2}\}^{m_2}...\{Z_{t_n}\}^{m_n}\right]=E\left[\{Z_{t_1+k}\}^{m_1}\{Z_{t_2+k}\}^{m_2}...\{Z_{t_n+k}\}^{m_n}\right]$$

for any k, and all integers $m_i > 0$ for i=1,2,...,n such that $\sum_{i=1}^{n} m_i \leq m$.

$\{Z_t\}$ is stationary up to order 1 if

$$E[Z_t]=\mu$$

where μ is a constant independent of time t.

$\{Z_t\}$ is stationary up to order 2 if;

(i) $E[Z_t]=\mu$

(ii) $E[Z_t^2]=\mu_2'$ which is a constant independent of time *t*. Hence

$$\text{var}[Z_t]=E[(Z_t-\mu)^2]=\mu_2'-\mu^2=\sigma^2$$

is also a constant, independent of time *t*.

(iii) $E[Z_t^{m_1}Z_s^{m_2}]=E[Z_0^{m_1}Z_{s-t}^{m_2}]$, by taking k=-t, which is a function of (s-t) only.

Hence

$$\text{cov}\{Z_t,Z_s\}=E[(Z_t-\mu)(Z_s-\mu)]=E[Z_tZ_s]-\mu^2$$

is also a function of (s-t) only.

Therefore, the mean and variance of Z_t for a second order stationary series remain constant over time, with the covariance between any two variables, Z_t and Z_s, depending only on the separation between the time points and not on their individual locations.

However the more moments that we specify the more information we have about a distribution. It is well accepted that processes with stationarity of order 2, known as weak stationarity are highly similar to processes that are stationary with higher orders. However Gaussian processes which are stationary up to order 2 are completely stationary.

2.2.2 Autocovariance and Autocorrelation Functions

Assuming the series $\{Z_t\}$, $t=0,\pm1,\pm2,\ldots$, is stationary up to order 2, we note from above that

$$\operatorname{cov}\{Z_t, Z_s\} = E\left[(Z_t - \mu)(Z_{t+k} - \mu)\right]$$

with the covariance between Z_t and Z_{t+k} depending only on k and not on t.

Then,

$$R(k) = E\left[(Z_t - \mu)(Z_{t+k} - \mu)\right],\ k=0,\pm1,\pm2,\ldots$$

where $R(k)$ is called the autocovariance function of $\{Z_t\}$ which measures the covariance between pairs of values of the series separated by an interval k, known as the lag. For $k=0$, we note that $R(0)$ is simply the variance of $\{Z_t\}$.

Also,

$$\rho(k) = \frac{R(k)}{R(0)},\ k=0,\pm1,\pm2,\ldots$$

where $\rho(k)$ is known as the autocorrelation function of $\{Z_t\}$. This is the correlation coefficient between the variables Z_t and Z_{t+k}. Thus $\rho(0) = \frac{R(0)}{R(0)} = 1$.

2.3 Purely Random Processes

A series $\{Z_t\}$, t=0,±1,±2,..., is called a purely random process or "white noise" if the series is a sequence of uncorrelated random variables denoted by $\{\varepsilon_t\}$. Therefore, since $\{Z_t\}$ is uncorrelated with all past and future values the series can be said as having "no memory". Certain series that consists simply of a sequence of independent random variables denoted by $\{e_t\}$ are those generated by a "strict white noise" process. If this series is second-order stationary, then for all t

$$E[Z_t]=\mu,$$
$$\text{var}[Z_t]=\sigma^2$$

and the autocovariance function is only a function of k since

$$R(k)=\text{cov}\{Z_t,Z_s\}=\begin{cases}\sigma^2 & \text{if } k=0\\ 0 & \text{if } k=\pm1,\pm2,....\end{cases}$$

Hence, the autocorrelation function is therefore

$$\rho(k)=\begin{cases}1 & \text{if } k=0\\ 0 & \text{if } k=\pm1,\pm2,....\end{cases}$$

2.4 Gaussian Processes

The series $\{Z_t\}$, t=0,±1,±2,..., is called a Gaussian process if, for any set of times $t_1,t_2,...,t_n$, the joint probability distribution of $\{Z_{t_1},Z_{t_2},...,Z_{t_n}\}$ is multivariate normal. Since a multivariate normal distribution is completely determined by its means, variances and covariances, the probabilistic structure of a Gaussian process is

completely determined by its "mean value function" $E[Z_t]$, and its "covariance function" $\text{cov}(Z_t, Z_s)$. It follows that, if a Gaussian process is second-order stationary so that the mean and covariance functions are invariant under a time shift, then the multivariate distribution of $\{Z_{t_1}, Z_{t_2}, ..., Z_{t_n}\}$ is also invariant under a shift in time and so the process is completely stationary. If the series is a Gaussian process then the sequence of uncorrelated random variables, $\{\varepsilon_t\}$ is also Gaussian and due to the zero correlation between values at different time points the process must now also be independent. Hence, the distinction between the sequence of random variables $\{\varepsilon_t\}$ and $\{e_t\}$ disappears.

2.5 Linear Time Series Models

Although the literature documents many definitions of a linear time series model, we refer to the simplest, that states that 'a linear time series is one in which an observation at time *t* can be predicted by a linear combination of past values', (Hannan (1986) and Hannan and Heyde (1972)).

It is said to have begun with Yule's Autoregressive (AR) model (1927), introduced for the study of sunspot numbers. An autoregressive model of order *p* denoted by AR(*p*) takes on the form

$$Z_t = \sum_{j=1}^{p} a_j Z_{t-j} + e_t \qquad (2.1)$$

where $a_1, a_2, ..., a_p$ are real constants, and $\{e_t\}$ is a purely random process. These form a significant class which have been found to provide useful descriptions for random processes arising in a large variety of problems. However, there is another general class of models known as moving average processes which also frequently occur in practical

applications. We say that Z_t is a moving average model of order q denoted by MA(q) if it satisfies an equation of the form

$$Z_t = \sum_{j=0}^{q} b_j e_{t-j} \tag{2.2}$$

where $b_0, b_1, ..., b_q$ are real constants. Also, without any loss of generality, we may assume that $b_0 = 1$. Since both AR and MA models have been found to provide useful descriptions for many real life processes, the two models were combined to construct a more general model. This was hence named an Autoregressive Moving Average (ARMA) model. We say that Z_t is an autoregressive moving average model of order (p,q) denoted by ARMA(p,q) if it can be expressed in the form

$$Z_t = \sum_{j=1}^{p} a_j Z_{t-j} + \sum_{j=0}^{q} b_j e_{t-j} \, . \tag{2.3}$$

The ARMA model includes both the AR and MA models as special cases. This is obtained either by setting $b_1 = b_2 = ... = b_q = 0$ or by setting $a_1 = a_2 = ... = a_p = 0$ respectively.

2.6 Maximum Likelihood Estimation Method

The three commonly used estimation procedures that guide us towards the construction of a suitable estimator are maximum likelihood, least squares and methods of moments. However, the former method as developed by Fisher (1922) is by far the most general and powerful method of estimation, which may be applied to any type of estimation problem provided that we can write down the joint probability distribution of the observations. The basic principles underlying this method of estimation are explained as follows:

Suppose we have a random sample of n independent observations given by $Z=(Z_1, Z_2, \ldots, Z_n)$ from a distribution with p.d.f. $f(z,\theta)$ involving a vector of unknown parameters $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$. The likelihood function of θ is the joint density of Z, which may be written as

$$l(\theta|z) = \prod_{i=1}^{n} f(z_i, \theta).$$

The method of maximum likelihood is based on the principle that we should choose the estimators $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)$ of θ, which maximises the likelihood function $l(\theta|z)$. However, it is usually more convenient to determine $\hat{\theta}$ by maximising the logarithm of the likelihood function $L(\theta) = \ln l(\theta|z)$, rather than the likelihood function itself.

2.7 Parameter Estimation for Time Series with Missing Values

It is very often the case in practice that observations corresponding to certain time points fail to be recorded or are recorded with gross errors. The Gaussian likelihood of time series with missing values can be calculated by application of the innovations algorithm to the theoretical covariance matrix of the observed values. However, the algebra involved is extremely complicated and errors in the computation of numerical operations appear. A method developed by Jones (1980) evaluates the exact Gaussian likelihood function of irregularly spaced observations from an invertible ARMA process. This is achieved by artificially augmenting the number of observations to a complete time series with independent standard normal random variables, which are independent of the observed values. Below, we describe this method by considering the invertible ARMA(p,q) process

$$Z_t - a_1 Z_{t-1} - \ldots - a_p Z_{t-p} = e_t + b_1 e_{t-1} + \ldots + b_q e_{t-q}$$

where the vectors a,b are real constants and $\{e_t\} \sim N(0,\sigma^2)$. The n independent observations are given by the vector $\boldsymbol{Z}_n = (Z_1, Z_2, \ldots, Z_n)'$ where the r observed values are $\boldsymbol{Z} = (Z_{j1}, Z_{j2}, \ldots, Z_{jr})'$ and the $m=n-r$ missing values are $\boldsymbol{Z}^m = (Z_{i1}, Z_{i2}, \ldots, Z_{im})'$. We artificially augment the number of observations in $\boldsymbol{Z}$ to n by the following method.

Here, we define

$$\boldsymbol{X}_{t+1} = F\boldsymbol{X}_t + T e_{t+1}, \tag{2.4}$$

$$\boldsymbol{Z}_t = G\boldsymbol{X}_t, \; t=1,2,3,\ldots \tag{2.5}$$

and

$$\boldsymbol{Y}_t = G_t\boldsymbol{X}_t + \alpha_t W_t, \; t=1,2,3,\ldots \tag{2.6}$$

where $F = \left[a_1\delta_{i,j} + \delta_{i+1,j}\right]_{i,j=1}^{r}$, $T = [1, b_1, \ldots, b_{r-1}]'$, $G = [1,0,\ldots,0]$ for $r = \max(p, q+1)$ and $\{W_t\}$ is an independent standard normal sequence, independent of $\{X_1, e_2, e_3, \ldots\}$.

Also,

$$\begin{cases} G_t = 0 \text{ and } \alpha_t = 1 & \text{if } Z_t \text{ is missing} \\ G_t = G \text{ and } \alpha_t = 0 & \text{if } Z_t \text{ is not missing.} \end{cases} \tag{2.7}$$

The vector $\boldsymbol{Y}_n = (Y_1, Y_2, \ldots, Y_n)'$ then coincides with $\boldsymbol{Z}_n = (Z_1, Z_2, \ldots, Z_n)'$ except for the unobserved components of $\boldsymbol{Z}_n$ which are replaced in $\boldsymbol{Y}_n$ by the independent standard normal random variables.

The Gaussian likelihood of the observed values $\boldsymbol{Z}$ is expressed in form

$$L\left(a,b,\sigma^2;\boldsymbol{Z}\right)=\left(\prod{}^{*}2\pi s_i\right)^{-1/2}\exp\left(-\frac{1}{2}\sum{}^{*}\frac{\left(Z_i-\hat{Y}_i\right)^2}{s_i}\right) \tag{2.8}$$

where $\hat{Y}_k$, k=1,2,...n are the one-step predictors and s_k, k=1,2,...n are their mean squared errors. Also, $\prod^{*}$ and $\sum^{*}$ denote the product and sum over only those indices i for which Z_i is observed.

A simple example is now given to further explain the above method of evaluating the Gaussian likelihood function for irregularly spaced observations from an invertible ARMA process.

Example: Consider the invertible ARMA(1,0) process

$$Z_t=aZ_{t-1}+e_t$$

where a is a real constant and $\{e_t\}\sim N\left(0,\sigma^2\right)$. We wish to calculate the Gaussian likelihood of the observed values $\boldsymbol{Z}=\left(Z_1,Z_4,Z_5\right)'$.

Equations (2.4) to (2.7) become

$$X_{t+1}=aX_t+e_{t+1},$$

$$Z_t=X_t$$

and

$$Y_t=G_tX_t+\alpha_tW_t$$

where $F=a$, $G=1$ and $T=1$.

Also,

$$\begin{cases} G_i = 0 \text{ and } \alpha_i = 1 & \text{for } i = 2,3 \\ G_i = 1 \text{ and } \alpha_i = 0 & \text{for } i = 1,4,5. \end{cases}$$

Now the extended observation vector is $Y_5=(Z_1, W_2, W_3, Z_4, Z_5)'$. The one-step predictors $\hat{Y}_k$, and their mean squared errors s_k, for k=1,2,...,5 are

$$\hat{Y}_1=0 \qquad \hat{Y}_2=0 \qquad \hat{Y}_3=0 \qquad \hat{Y}_4=a^3 Z_1 \qquad \hat{Y}_5=aZ_4$$

and

$$s_1=\frac{\sigma^2}{1-a^2} \qquad s_2=1 \qquad s_3=1 \qquad s_4=\sigma^2\left(1+a^2+a^4\right) \qquad s_5=\sigma^2.$$

Therefore, the Gaussian likelihood of the observed values $\boldsymbol{Z}$, given by equation (2.8) is

$$L\left(a,\sigma^2;\boldsymbol{Z}\right)=\sigma^{-3}(2\pi)^{-3/2}\left(\frac{1+a^2+a^4}{1-a^2}\right)^{-1/2}$$
$$\exp\left(-\frac{1}{2\sigma^2}\left[Z_1^2\left(1-a^2\right)+\frac{\left(Z_4-a^3Z_1\right)^2}{\left(1+a^2+a^4\right)}+\left(Z_5-aZ_4\right)^2\right]\right).$$

Chapter Three

Non-Linear Time Series

Since most aspects of the real world are concerned with non-linear relationships between variables and therefore non-linear processes, many authors such as Tong (1978, 1983), Granger and Anderson (1978), Priestly (1980), Tong and Lim (1980) and Haggen and Ozaki (1981) argued for the development of non-linear time series.

In recent years, we have witnessed an ever-increasing interest in non-linear time series analysis, which is now considered a highly developing subject. These models provide a more accurate description of the real world. Tong (1990) demonstrates this by providing examples of data sets including animal populations, medical science and economics.

In this chapter we will discuss some special non-linear models that have been successfully applied to a wide range of real data.

3.1 Non-Linear Models

3.1.1 Bilinear Model

These time series models were first introduced in the control theory literature. This can be seen for example, in Ruberti et al. (1972), Mohler (1973), and Brockett (1976). These studies consider continuous time models and describe the relationship between observable input and output processes of physical systems.

However, the discrete time bilinear model denoted by BL takes the general form

$$Z_t + \sum_{j=1}^{p} a_j Z_{t-j} = \sum_{i=1}^{m}\sum_{j=1}^{n} b_{ij} Z_{t-i} e_{t-j} + \sum_{j=0}^{q} c_j e_{t-j} \qquad (3.1)$$

where a, b are real constants, $c_0=1$ and $\{e_t\}$ is a strict white noise process. It is easily seen that equation (3.1) is an enlargement of the ARMA model obtained by adding a BL form in $\{Z_{t-i}, e_{t-j}\}$ to the R.H.S. of equation (2.3). If we set $b_{ij}=0$ for all i,j, then the general BL model can clearly be seen to include the standard ARMA models as a special case.

3.1.2 Threshold Autoregressive (TAR) Models

This class of non-linear model was first introduced by Tong (1978), for the analysis of river flow data. To explain, we start with a linear model for a time series $\{Z_t\}$ and then

allow the parameters of the model to vary according to the values of a finite number of past values. We say that Z_t is a first-order two régime threshold autoregressive model denoted by TAR(1), if it can be expressed in the form

$$Z_t = \begin{cases} a^{(1)} Z_{t-1} + e_t^{(1)} & \text{if } Z_{t-1} < d \\ a^{(2)} Z_{t-1} + e_t^{(2)} & \text{if } Z_{t-1} \geq d \end{cases}$$

where $a^{(1)}$, $a^{(2)}$, are real constants, $\{e_t^{(1)}\}$, $\{e_t^{(2)}\}$ are each strict white noise processes and d is a constant known as the threshold. This model for example, can be readily extended to a first-order l-régime model, which would typically be of the form

$$Z_t = a^{(i)} Z_{t-1} + e_t^{(i)} \quad \text{if } Z_{t-1} \in (r_{i-1}, r_i] = \boldsymbol{R}_i,\ i = 1,2,\ldots,l$$

where $-\infty = r_0 < r_1 < \ldots < r_l = \infty$ are the threshold values and so $\boldsymbol{R}_1, \boldsymbol{R}_2, \ldots, \boldsymbol{R}_l$ are a given partition of the real line $\boldsymbol{R}$. Hence this model may be regarded as 'piecewise linear' approximation to the general non-linear AR(1) model

$$Z_t = \lambda(Z_{t-1}) + e_t$$

where $\lambda(.)$ is some general non-linear scalar valued function of a single real variable.

A tenuous link between BL and TAR models is shown (Priestley (1988)), by considering a first-order bilinear system in which the input U_t, output Z_t, and noise e_t are related by

$$Z_{t+1} = aZ_t + cU_tZ_t + e_{t+1}.$$

If

$$U_t = \begin{cases} \alpha & \text{if } Z_t < d \\ -\alpha & \text{if } Z_t \geq d \end{cases}$$

then the model for Z_t becomes

$$Z_{t+1}=\begin{cases} a^{(1)}Z_t + e_{t+1} & \text{if } Z_t < d \\ a^{(2)}Z_t + e_{t+1} & \text{if } Z_t \geq d \end{cases}$$

where $a^{(1)}=a+\alpha c$ and $a^{(2)}=a-\alpha c$.

We say that Z_t is an l-régime threshold autoregressive model of order p denoted by TAR(p) if it can be expressed in the form

$$Z_t=a_0^{(i)} + \sum_{j=1}^{p} a_j^{(i)} Z_{t-j} + e_t^{(i)} \quad \text{if } Z_{t-d} \in \boldsymbol{R}_i,\ i=1,2,\ldots,l$$

where $\boldsymbol{R}_i$ is a given subset of the real line $\boldsymbol{R}$, the $a_0^{(i)}$'s are a set of l real constants and the $e_t^{(i)}$'s are an independent sequence of zero mean i.i.d. random variables, having a finite variance σ^2.

However, as discussed later we are concerned with an l-régime zero-order threshold autoregressive model, which is given by

$$Z_t=a^{(i)} + e_t^{(i)} \quad \text{if } Z_{t-1} \in \boldsymbol{R}_i,\ i=1,2,\ldots,l. \tag{3.2}$$

3.2 Parameter Estimation for Non-Linear Time Series Models

Although approaches for the estimation of non-linear time series remain similar to that of linear time series, (i.e. maximum likelihood estimation), relatively little success has so far been achieved in the estimation of non-linear time series. This is partially due to the vast range of possible models.

However as forecasting is one of the main purposes of time series analysis, Granger and Anderson (1978) only consider models with the ability of producing forecasts. Petruccelli and Davies (1986) develop a test for estimating TAR(k) models, which was then modified by Pemberton (1993, 1994).

Chapter Four

Mixture Models

Mixture models were first introduced by Jalali and Pemberton (1995). In their study, they considered two types of mixture models, firstly the Bernstein mixture model, which uses Bernstein polynomials and the Laguerre mixture model which uses Laguerre polynomials. These models have a much richer structure than zero-order threshold models and contain the latter as a special case. Further, these models have the same autocovariance structure as that of linear ARMA models, although with these models only a subset of the possible autocovariance functions with ARMA models are obtained.

Nevertheless, they have the important property of duality, which corresponds to time reversal and so mixture models are at least as exciting as linear models. Jalali and Pemberton (1995) explained mixture models for time series in terms of probability measures. However, in this chapter we discuss the underlying ideas and properties of mixture model theory in terms of probability density functions.

4.1 The Basic Model

Let $f=(f_1,f_2,...,f_k)$ be a vector of linearly independent p.d.f.'s each defined on a sample space S, where we assume for simplicity, S is the real line, R.

Denote by U the random vector $(U_1,U_2,...,U_k)$ where the i^{th} element U_i has density f_i. Similarly $U^{(t)}$ for t=1,2,... is an i.i.d. sequence of replicas of U at time t.

Define $W=\left\{w=(w_1,w_2,...,w_k):\sum_{i=1}^{k} w_i=1 \;\&\; w_i\geq 0\ \forall i\right\}$ then by a w-mixture of U we imply the random variable whose p.d.f. is $\sum_{i=1}^{k} w_i f_i$.

Proposition 4.1.1: For every mixture of the vector of linearly independent densities $f=(f_1,f_2,...,f_k)$, a unique vector of weights in w is associated.

Proof

Suppose the random vector V has two representations as a mixture of U, with coefficients $q=(q_1,q_2,...,q_k)$, $r=(r_1,r_2,...,r_k)$ and that for some i, $q_i\neq r_i$. Then the p.d.f. of V can be written in two ways,

$$\sum_{i=1}^{k} q_i f_i = \sum_{i=1}^{k} r_i f_i \text{ which implies } \sum_{i=1}^{k} (q_i - r_i) f_i = 0.$$

The definition of linear independence states that if a given set of scalars $c=(c_1, c_2, \ldots, c_k)$ has p.d.f. $\sum_{i=1}^{k} c_i f_i(.) \equiv 0$ implies $c \equiv 0$. Therefore, since the f_i's are linearly independent, $(q_i - r_i)=0$ for all i's. This contradicts the previous assumption that $q_i \neq r_i$ for some i. (Q.E.D.)

Let $\alpha : R \to W$ be defined as $\alpha(s)=(\alpha_1(s), \alpha_2(s), \ldots, \alpha_k(s))$. Of course, this means $\alpha_i(s) \geq 0$ and $\sum_{i=1}^{k} \alpha_i(s)=1$ for all $s \in R$. These are the unique vector of weights which builds up the dependence structure from the sequence $U^{(t)}$.

To determine a time series $(Z_t : t = 0,1,2,\ldots)$ all that is required is Z_0, α and f.

Let Z_0 be the initial element and successively take Z_t as the $\alpha(Z_{t-1})$ mixture of $U^{(t)}$, for t=1,2,....

Hence the p.d.f. of Z_t given $Z_{t-1}, Z_{t-2}, \ldots, Z_0$ is

$$\sum_{i=1}^{k} \alpha_i(z_{t-1}) f_i(z_t). \qquad (4.1)$$

4.2 Zero-Order Threshold Autoregressive Model as a Special Case of Mixture Models

As seen in equation (3.2), the zero-order k-régime threshold model takes the general form

$$Z_t = a^{(i)} + e_t^{(i)}, \quad \text{if } Z_{t-1} \in \boldsymbol{R}_i, \; i=1,2,...,k.$$

The $a^{(i)}$'s are a set of k real constants and the $e_t^{(i)}$'s are the independent sequences of i.i.d. random variables with densities g_i, say, with mean zero and finite variance.

Let $\boldsymbol{R}_1, \boldsymbol{R}_2, ..., \boldsymbol{R}_k$ be a measurable partition of the real line, $\boldsymbol{R}$, with α_i being the indicator function of $\boldsymbol{R}_i$, i.e.

$$\alpha_i(z) = \begin{cases} 1 & \text{if } z \in \boldsymbol{R}_i \\ 0 & \text{if } z \notin \boldsymbol{R}_i \end{cases}$$

Setting the i^{th} element of the random vector $\boldsymbol{U}=(U_1, U_2, ..., U_k)$ at time t to the zero-order k-régime threshold model given above i.e. $U_i^{(t)} = a^{(i)} + e_t^{(i)}$, then the U_i's have p.d.f.s $f_{U_i}(u_i) = g_i\left(u_i - a^{(i)}\right)$.

Clearly, the conditional p.d.f. of Z_t given Z_{t-1} is then

$$f_{Z_t|Z_{t-1}}(z_t|z_{t-1}) = \sum_{i=1}^{k} \alpha_i(z_{t-1}) g_i\left(z_t - a^{(i)}\right). \tag{4.2}$$

This is the $\boldsymbol{\alpha}(z_{t-1})$ mixture of $\boldsymbol{U}^{(t)}$ and so the k-régime threshold model of order zero is shown to be nothing more nor less than a mixture model.

4.3 Special Properties of Mixture Models

The conditional distribution of Z_t given $Z_{t-1}, Z_{t-2}, \ldots, Z_0$ is shown to have in equation (4.1) the p.d.f. $\sum_{i=1}^{k} \alpha_i(z_{t-1}) f_i(z_t)$. The conditional distribution of Z_t is thus clearly dependent only on Z_{t-1} and so the time series $\{Z_t : t = 0,1,2,\ldots\}$ is Markovian.

Hence,

$$f_{Z_t|Z_{t-1},Z_{t-2},\ldots,Z_0}(z_t|z_{t-1}, z_{t-2}, \ldots, z_0) = f_{Z_t|Z_{t-1}}(z_t|z_{t-1}). \tag{4.3}$$

Proposition 4.3.1: The elements of the time series $\{Z_t : t = 0,1,2,\ldots\}$, apart from the initial one, Z_0, are all mixtures of U.

Proof

$$\begin{aligned}
f_{Z_t}(z_t) &= \int_{-\infty}^{\infty} f_{Z_t,Z_{t-1}}(z_t, z_{t-1}) dz_{t-1} \\
&= \int_{-\infty}^{\infty} f_{Z_t|Z_{t-1}}(z_t|z_{t-1}) f_{Z_{t-1}}(z_{t-1}) dz_{t-1} \\
&= \int_{-\infty}^{\infty} \sum_{i=1}^{k} \alpha_i(z_{t-1}) f_i(z_t) f_{Z_{t-1}}(z_{t-1}) dz_{t-1} \\
&= \sum_{i=1}^{k} \left(\int_{-\infty}^{\infty} \alpha_i(z_{t-1}) f_{Z_{t-1}}(z_{t-1}) dz_{t-1} \right) f_i(z_t) \\
&= \sum_{i=1}^{k} a_i f_i(z_t)
\end{aligned}$$

Hence, we now need to show $a=(a_1, a_2, \ldots, a_k) \in W$.

$$\sum_{i=1}^{k} a_i = \sum_{i=1}^{k} \left(\int_{-\infty}^{\infty} \alpha_i(z_{t-1}) f_{Z_{t-1}}(z_{t-1}) dz_{t-1} \right)$$

$$= \int_{-\infty}^{\infty} \left(\sum_{i=1}^{k} \alpha_i(z_{t-1}) \right) f_{Z_{t-1}}(z_{t-1}) dz_{t-1}$$

$$= \int_{-\infty}^{\infty} f_{Z_{t-1}}(z_{t-1}) dz_{t-1} \text{ since } \sum_{i=1}^{k} \alpha_i(z_{t-1}) = 1$$

$$= 1$$

Clearly, since $\alpha_i(.) \geq 0$ and $f_{Z_{t-1}}(.) \geq 0, a_i \geq 0$. (Q.E.D.)

Now define

$$p_{ij} = \int_{-\infty}^{\infty} \alpha_j(z) f_i(z) dz \tag{4.4}$$

and let $\boldsymbol{P}$ be the transition matrix of the Markov chain whose $(i,j)^{\text{th}}$ entry is p_{ij}.

Proposition 4.3.2: $\boldsymbol{P}$ is a stochastic matrix i.e. the entries are ≥ 0 and the row sums are 1.

Proof

The entries p_{ij} of the square $k \times k$ matrix $\boldsymbol{P}$ are all ≥ 0 since the α_j's are weights and the f_i's are the probability densities. Now, for any i, $1 \leq i \leq k$,

$$\sum_{j=1}^{k} p_{ij} = \sum_{j=1}^{k}\left(\int_{-\infty}^{\infty}\alpha_j(z)f_i(z)dz\right)$$

$$= \int_{-\infty}^{\infty}\left(\sum_{j=1}^{k}\alpha_j(z)\right)f_i(z)dz$$

$$= \int_{-\infty}^{\infty} f_i(z)dz \text{ since } \sum_{j=1}^{k}\alpha_j(z)=1$$

$$=1. \qquad \text{(Q.E.D.)}$$

From proposition 4.3.1, we know that Z_t is a p_t mixture of $U^{(t)}$. This result gives us the following proposition.

Proposition 4.3.3: The vectors $p_1, p_2, \ldots, p_t, \ldots$ constitute the consecutive probability vectors of a k-state homogeneous Markov chain with transition matrix P where $(p_1)_i = \int_{-\infty}^{\infty}\alpha_i(z_0)f_{Z_0}(z_0)dz_0$ is the i^{th} component of p_1. Then at time t the following relationship holds

$$p_t = p_{t-1}P. \tag{4.5}$$

Proof

From the above proof of proposition 4.3.1

$$(p_t)_i = \int_{-\infty}^{\infty}\alpha_i(z_{t-1})f_{Z_{t-1}}(z_{t-1})dz_{t-1}$$

$$= \int_{-\infty}^{\infty}\alpha_i(z_{t-1})\sum_{j=1}^{k}(p_{t-1})_j f_j(z_{t-1})dz_{t-1}$$

$$= \sum_{j=1}^{k} (p_{t-1})_j \int_{-\infty}^{\infty} \alpha_i(z_{t-1}) f_j(z_{t-1}) dz_{t-1}$$

$$= \sum_{j=1}^{k} (p_{t-1})_j p_{ji}$$

$$= (p_{t-1}P)_i .$$

Hence $p_t = p_{t-1}P$. (Q.E.D.)

As a corollary to equation (4.5) above, the conditional distribution of Z_{t+1} given $Z_0 = z_0$ is

$$f_{Z_{t+1}|Z_0}(z_{t+1}|z_0) = \alpha(z_0)P^t f_{t+1}(z_{t+1}), \quad \text{for } t \geq 1.$$

We now assume that the transition matrix P is ergodic, which implies the corresponding finite state Markov chain is irreducible and aperiodic. Below, five preliminary definitions for the conditions of ergodicity are stated:

(i) If state j can be reached from state i in exactly n-steps then state j is said as being accessible from state i with transition probability denoted by $p_{ij}^{(n)}$.

(ii) States i and j communicate if state j is accessible from state i and vice-versa.

(iii) State i is an absorbing state if once entered it cannot be left (i.e. the transition probabilities $p_{ii}=1$).

(iv) A Markov chain is irreducible if, and only if, all pairs of states communicate. Therefore, if there is an absorbing state the Markov chain is not irreducible.

(v) If return to a state i is possible in n-steps, when n is a multiple of an integer t and t is the greatest such integer, then state i is said to have period t.

If t=1 then state i is aperiodic.

If t>1 then state i is periodic.

A state i to which no return is possible (i.e. the transition probability $p_{ii}^{(n)}=0$ for all $n>0$) will be considered aperiodic.

It follows that $\boldsymbol{P}$ converges to a matrix with equilibrium state $\boldsymbol{p}_\infty$ denoted by π, where its components are $\pi_1,\pi_2,\ldots,\pi_k$. This is the eigenrow of the matrix $\boldsymbol{P}$ with the eigenvalue 1. Note that $\pi_j \neq 0$ for all j=1,2,...,k by ergodicity.

Proposition 4.3.4: If the transition matrix $\boldsymbol{P}$ is ergodic with equilibrium state $\pi=(\pi_1,\pi_2,\ldots,\pi_k)$, then the limiting stationary distribution of Z_{t+1} has p.d.f. $\sum_{j=1}^{k}\pi_j f_j(z_\infty)$.

Proof

We know, by hypothesis, that as $t\to\infty$, $\boldsymbol{P}^t \to \boldsymbol{\Pi}=\begin{pmatrix}\pi_1,\pi_2,\ldots,\pi_k\\ \pi_1,\pi_2,\ldots,\pi_k\\ \\ \\ \pi_1,\pi_2,\ldots,\pi_k\end{pmatrix}$

Therefore, from the previous corollary, as $t\to\infty$,

$$f_{Z_{t+1}|Z_0}(z_{t+1}|z_0)\to\boldsymbol{\alpha}(z_0)\boldsymbol{\Pi}\boldsymbol{f}_\infty(z_\infty)$$

$$=\pi\ f_\infty(z_\infty) \text{ since } \sum_{i=1}^{k}\alpha(z_0)=1\ \forall z_0\in\boldsymbol{R}$$

$$= \sum_{j=1}^{k} \pi_j f_j(z_\infty). \qquad \text{(Q.E.D.)}$$

From now on, we will assume the time series $\{Z_t\}$ to have reached its stationary state (or, equivalently, that Z_0 has been drawn from the equilibrium distribution $f_\infty(.)$).

Jalali and Pemberton (1995) go on to discuss duality and time reversibility, together with two examples of models. Whilst, not directly relevant to the present study, it is worth recording that the Bernstein model, and the Laguerre model we consider in chapter five are both stationary and time reversible. That is, the joint density of $(Z_t, Z_{t-1}, \ldots, Z_1)$ is identical to that of $(Z_{-t}, Z_{-t-1}, \ldots, Z_{-1})$. In particular, this implies a symmetry of the bivariate density of Z_t and Z_{t-1}:

$$f_{Z_t, Z_{t-1}}(x, y) = f_{Z_{t-1}, Z_t}(x, y).$$

In the generality of mixture models, Jalali and Pemberton (1995) discuss the basic idea that a so-called dual can be constructed for any mixture model and the class of models is closed under duality. This means the dual is also a mixture model with weights α's and p.d.f.s f's. It has further been argued that the dual is the time reversal of the original model. Although, while a process which is self-dual (i.e. models which retain the same form for the α's and f's) is then time reversible, it does not follow that a time reversible process is self-dual due to the α's and f's of the dual possibly being different from those of the original process. A fuller discussion of duality is not relevant to the present work and we refer the reader to Jalali and Pemberton (1995).

Chapter Five

The Laguerre Mixture Model

In this chapter, we introduce the Laguerre mixture model, which is a fairly straightforward extension of a model with a finite number of mixture elements, to one with a countable number of mixture elements. In Beet (1995), a one-step conditional distribution for the Laguerre model was used for the likelihood function and subsequent estimation of the unknown parameters, since it was assumed that the time series data was complete. However, we wish to extend this procedure to a time series that has some missing values and so $\tau\,(\geq 1)$-step conditional distributions are required, where τ is a generic time lag between the observed values.

Further, the likelihood function is investigated using plots and numerical optimisation routines. Following this, we examine the sampling properties of the maximum likelihood estimates using Monte Carlo techniques for various sample sizes' and true values of the parameters.

5.1 The Laguerre Mixture Model

For the Laguerre model we allow the parameter k of the general mixture model to be countably large and set

$$\alpha_i(z)=(\theta z)^i \exp(-\theta z)/i! \qquad i=0,1,2,...;\ z\geq 0.$$
$$f_i(z)=\lambda(\lambda z)^i \exp(-\lambda z)/i! \qquad \theta>0,\ \lambda>0$$

The weights $\alpha_i(.)$ are thus the Poisson probabilities corresponding to mean θz and the density $f_i(.)$ is for the Gamma distribution with scale $1/\lambda$ and shape (i+1).

Here we note that, for the existence of a stationary distribution, it is necessary and sufficient that $\theta<\lambda$, hence $r=\theta/\lambda<1$. The stationary density, $f_\infty(z)$, for the Laguerre model is a mixture of the p.d.f.s $f_i(z)$ using the stationary weights $\pi_i=(1-r)r^i$ (derived in Appendix A). It then follows easily that f_∞ is an exponential distribution with mean $(\lambda(1-r))^{-1}$:

$$f_\infty(z)=\sum_{i=0}^{\infty}\pi_i f_i(z)$$
$$=\sum_{i=0}^{\infty}(1-r)r^i\lambda(\lambda z)^i\exp(-\lambda z)/i!$$

$$= \lambda(1-r)\exp(-\lambda z)\sum_{i=0}^{\infty}(\lambda z r)^i / i!$$

$$= \lambda(1-r)\exp(-\lambda z)\exp(\lambda z r)$$

$$= \lambda(1-r)\exp(-\lambda(1-r)z) \tag{5.1}$$

In Appendix A, we derive the conditional distribution of Z_t given the initial value Z_0 for the Laguerre model and show the density to be

$$f(z_t|z_0) = \lambda u_t \exp\left\{-\lambda u_t\left(z_t + z_0 r^t\right)\right\} I_0\left\{2\lambda u_t\sqrt{\left(z_0 z_t r^t\right)}\right\} \tag{5.2}$$

where $u_t = \dfrac{(1-r)}{\left(1-r^t\right)}$, and $I_0(.)$ is the modified Bessel function of the first kind of order zero. For order n, this is $I_n(x) = \left(\dfrac{x}{2}\right)^n \sum_{k=0}^{\infty}\left(\dfrac{x^2}{4}\right)^k \dfrac{1}{k!\Gamma(n+k+1)}$ hence

$$I_0(x) = \sum_{k=0}^{\infty}\left(\frac{x^2}{4}\right)^k \frac{1}{(k!)^2}.$$

We also observe that the Laguerre model is time reversible; the joint density of (Z_0, Z_t) can easily be obtained;

$$f(z_0, z_t) = f(z_t|z_0)f(z_0)$$

$$= \lambda u_t \exp\left(-\lambda u_t\left(z_t + z_0 r^t\right)\right) I_0\left(2\lambda u_t\sqrt{\left(z_0 z_t r^t\right)}\right)\lambda(1-r)\exp(-\lambda(1-r)z_0)$$

$$= \lambda^2(1-r)u_t \exp\left(-\lambda u_t r^t z_0 - \lambda(1-r)z_0\right)\exp(-\lambda u_t z_t)\ I_0\left(2\lambda u_t\sqrt{\left(z_0 z_t r^t\right)}\right)$$

$$= \lambda^2(1-r)u_t \exp\left(-\lambda u_t r^t z_0 - \lambda u_t\left(1-r^t\right)z_0\right)\exp(-\lambda u_t z_t)\ I_0\left(2\lambda u_t\sqrt{\left(z_0 z_t r^t\right)}\right)$$

$$= \lambda^2(1-r)u_t \exp(-\lambda u_t z_0)\exp(-\lambda u_t z_t)\ I_0\left(2\lambda u_t\sqrt{\left(z_0 z_t r^t\right)}\right).$$

$$= \lambda^2 (1-r) u_t \exp\left(-\lambda u_t (z_0 + z_t)\right) I_0\left(2\lambda u_t \sqrt{(z_0 z_t r^t)}\right) \tag{5.3}$$

and this is symmetric in (Z_0, Z_t). Hence the model is time reversible.

Figures 5.1 to 5.6 below, show graphs of the *t*-step conditional p.d.f.s given by equation (5.2) for values of $(r, \lambda) = (0.5, 10.0)$, t =1,5,100 and initial value z_0=1,5. In addition, the stationary density is shown, which as previously mentioned is an exponential distribution with mean $\left(\lambda(1-r)\right)^{-1}$. These plots indicate that for large *t*, the influence of the initial value z_0 gradually wears off and the *t*-step conditional p.d.f.s tends towards the exponential distribution with mean $\left(\lambda(1-r)\right)^{-1}$. However, as z_0 increases the exponential nature of this function takes longer to become apparent.

Plots of the *t*-step conditional p.d.f. and $\text{Exp}(\lambda(1-r))^{-1}$ with $(r,\lambda) = (0.5, 10.0)$

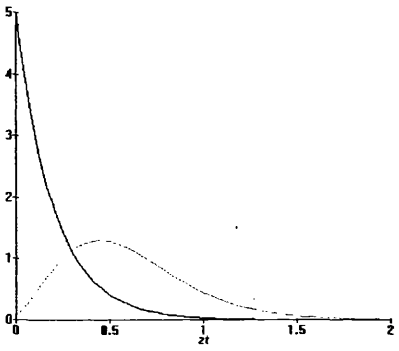


Figure 5.1: t=1.0, z_0=1.0

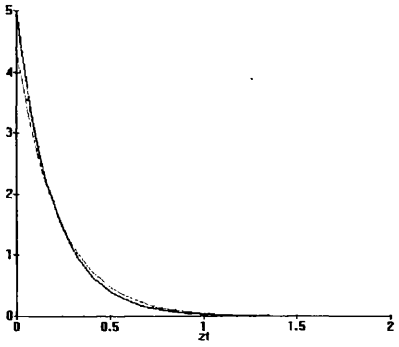


Figure 5.2: t=5.0, z_0=1.0

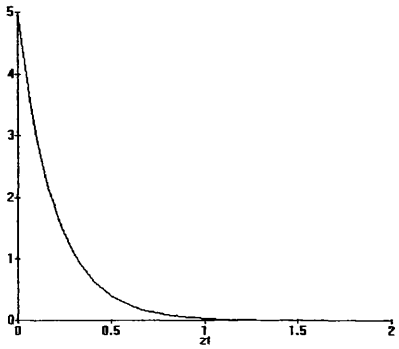


Figure 5.3: t=100.0, z_0=1.0

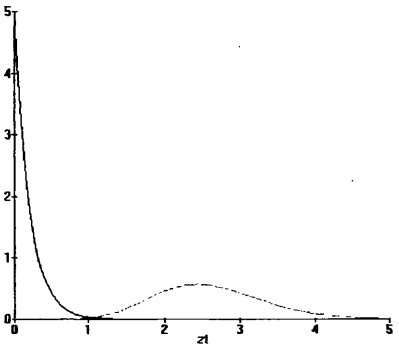


Figure 5.4: t=1.0, z_0=5.0

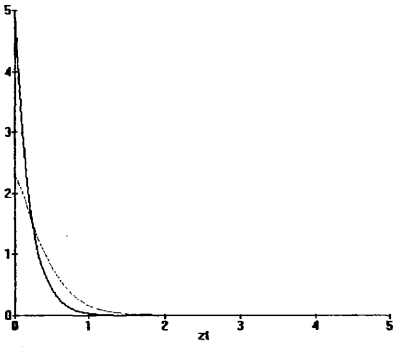


Figure 5.5: t=5.0, z_0=5.0

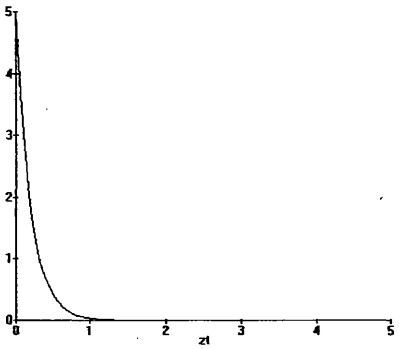


Figure 5.6: t=100.0, z_0=5.0

5.2 Modification of the Time Series to Create Zero's

Interest in non-negative time series arose when considering how to model river-flow, in particular periods without flow. Hutton (1990) explains, 'One mathematical formulation of a simple non-negative time series for streamflow starts from the conservation equation

$$S_{i+1}=S_i-Q_i+R_i$$

where S_i is the volume of water in the reservoir, Q_i is the volume of water flowing out and R_i is the input, all on day i. As we are working in discrete time, we approximate by considering the inputs to be added at the end of the interval, after the flow is lost.

Assuming that flow is proportional to volume stored,

$$Q_i=\kappa S_i \text{ and } S_{i+1}=(1-\kappa)S_i+R_i$$

where $0<\kappa<1$. The marginal distribution of the R_i, assumed to be independent and identically distributed and independent of S_i, has the form

$$R_i=\begin{cases}0 & \text{with probability } \delta\\ \mathrm{Y_i} & \text{with probability } 1-\delta\end{cases}$$

where Y_i is a continuous, positive random variable, with probability density function $f_y(\theta)$.' Indeed, it may be possible to approximate Y_i by a Laguerre time series.

Given a time series $Z_0,Z_1,\ldots,Z_t$ of length t+1 generated from the Laguerre model with parameters r and λ, we substitute a certain proportion, δ, of values in the series $Z_1,Z_2,\ldots,Z_{t-1}$ for zero's to indicate that the observation is missing. We then let

$t_0, t_1, \ldots, t_n$ be times such that the observations $Z_{t_i} > 0$. This is achieved by the distribution of H_i, which depends on δ but is independant of Z_i and has the form

$$H_i = \begin{cases} 0 & \text{if } Z_i \text{ is missing with probability } \delta \\ 1 & \text{if } Z_i \text{ is observed with probability } 1-\delta \end{cases}$$

where,

$$f\left(h_0, h_1, \ldots, h_t, z_{t_0}, z_{t_1}, \ldots, z_{t_n} \middle| \theta, \delta\right) = \delta^{t-(n+1)}(1-\delta)^{n+1} f\left(z_{t_0}, z_{t_1}, \ldots, z_{t_n} \middle| \theta\right).$$

Clearly, δ can easily be estimated and so we only wish to examine the estimation of $\theta = (r, \lambda)$ by maximum likelihood.

5.3 Model Estimation by Maximum Likelihood

As mentioned in section 2.6, the likelihood of $\theta = (r, \lambda)$ is obtained as the joint probability density function of the $n+1$ observed values, $z = \left(z_{t_0}, z_{t_1}, \ldots, z_{t_n}\right)$. The t_i's are the discrete set of times where an observation is obtained and w.l.o.g. we set $t_0 = 0$ for convenience with $t_n = t$ being the length of the observation period. The likelihood is thus expressed as

$$l(\theta|z) = f\left(z_{t_0}, z_{t_1}, \ldots, z_{t_n} \middle| \theta\right) = \left[\prod_{i=1}^{n} f\left(z_{t_i} \middle| z_{t_{i-1}}\right)\right] f(z_0) \tag{5.4}$$

where $f\left(z_{t_i} \middle| z_{t_{i-1}}\right)$ is a $\tau_i \left(= t_i - t_{i-1}\right)$-step conditional density, and $f(z_0)$ is the stationary density. Substitution of the conditional density, (5.2) and the stationary density, (5.1) gives

$$l(\theta|z)=\prod_{i=1}^{n}\left[\lambda u_{\tau_i}\exp\left\{-\lambda u_{\tau_i}\left(z_{t_i}+z_{t_{i-1}}r^{\tau_i}\right)\right\}I_0\left\{2\lambda u_{\tau_i}\sqrt{\left(z_{t_{i-1}}z_{t_i}r^{\tau_i}\right)}\right\}\right]$$

$$\lambda(1-r)\exp\left\{-\lambda(1-r)z_0\right\}, \tag{5.5}$$

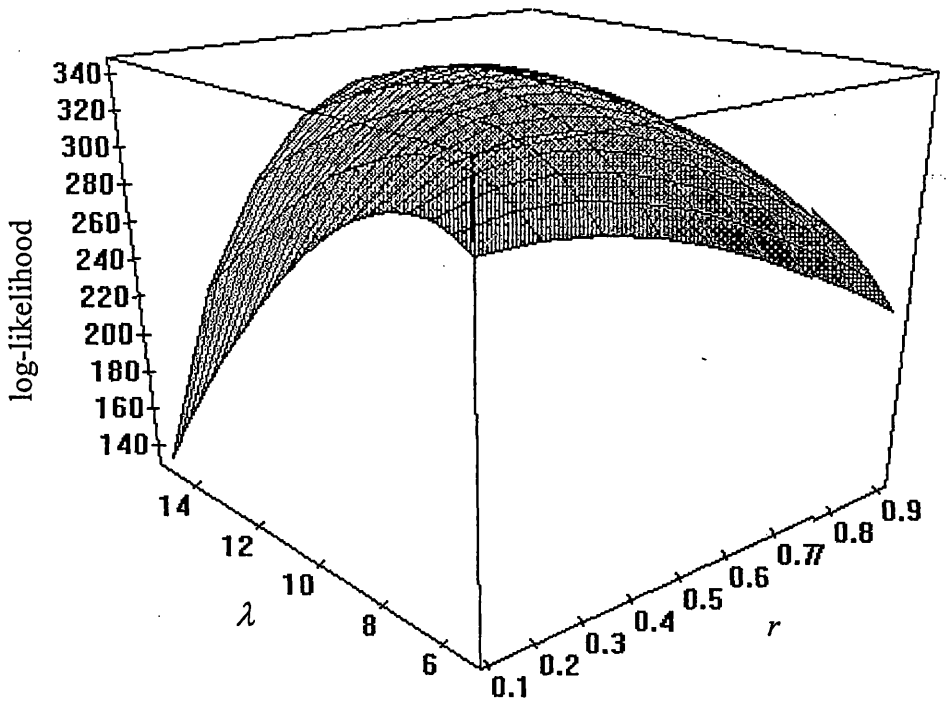
Taking natural logarithms gives the log-likelihood as a function of $\theta = (r, \lambda)$,

$$L(\theta)=\ln l(\theta|z) \quad =\sum_{i=1}^{n}\left[\ln u_{\tau_i}-\lambda u_{\tau_i}\left(z_{t_i}+z_{t_{i-1}}r^{\tau_i}\right)+\ln I_0\left(2\lambda u_{\tau_i}\sqrt{\left(z_{t_{i-1}}z_{t_i}r^{\tau_i}\right)}\right)\right]$$

$$+(n+1)\ln\lambda+\ln(1-r)-\lambda(1-r)z_0 \tag{5.6}$$

To give some idea of the appearance of this function, we used the Fortran 77 program as shown in Appendix B to simulate a time series of length t+1=500 observations from the Laguerre model with true values of $(r, \lambda) = (0.5, 10.0)$. However, to produce the data with missing values included, we decided to delete 10% of the observations at random for zero. The log-likelihood function is then computed over the range $r \in [0.1,0.9]$ and $\lambda \in [5.0,15.0]$. These results are shown in Figure (5.7) below.

Figure (5.7): Plot of Log-Likelihood Function.



The location of the maximum point (r_1, λ_1) is found by solving

$$\frac{\partial L(\theta)}{\partial r}=0 \text{ and } \frac{\partial L(\theta)}{\partial \lambda}=0 \tag{5.7}$$

subject to the determinant of the Hessian matrix, Δ, defined by

$$\Delta=\left\{\left(\frac{\partial^2 L}{\partial r^2}\right)\left(\frac{\partial^2 L}{\partial \lambda^2}\right)-\left(\frac{\partial^2 L}{\partial r \partial \lambda}\right)^2\right\}\Bigg|_{(r_1,\lambda_1)} > 0 \tag{5.8}$$

and

$$\left.\frac{\partial^2 L}{\partial \lambda^2}\right|_{(r_1,\lambda_1)} < 0 \left(\text{ or } \left.\frac{\partial^2 L}{\partial r^2}\right|_{(r_1,\lambda_1)} < 0\right). \tag{5.9}$$

However, due to the complexity of the first derivatives of $L(\theta)$ w.r.t. r and λ we have been unable to find a critical point analytically. Therefore, we can only investigate whether a maximum point does exist through plots and numerical optimisation routines. However Figure (5.7) suggests there is a unique maximum point, in this case not too far from the true values of r and λ. In addition, the results of the simulations for r and λ estimates, shown later, always provide us with a unique solution.

After some preliminaries, we demonstrate graphically and numerically that there seems to be a maximum point. This is indicated since in the vicinity of this point, Δ is greater than zero with $\frac{\partial^2 L}{\partial \lambda^2} < 0$. It follows that $\frac{\partial^2 L}{\partial r^2} < 0$.

The derivatives of the log-likelihood function $L(\theta)$ are relatively straightforward on noting that $I_0'(x) = I_1(x)$ which is a modified Bessel function of the first kind of order one, where its derivative $I_1'(x) = I_0(x) - \frac{1}{x} I_1(x)$.

Recalling that $\theta = (r, \lambda)$, let $B(\theta : x_i) = \frac{x_i I_1(x_i)}{I_0(x_i)}$, where $x_i = 2\lambda u_{\tau_i} \sqrt{\left(z_{t_{i-1}} z_{t_i} r^{\tau_i}\right)}$ for i=1,2,...,n.

The first derivatives of $L(\theta)$ are

$$\frac{\partial L(\theta)}{\partial r} = \sum_{i=1}^{n} \left[-\lambda u_{\tau_i} z_{t_{i-1}} \tau_i r^{\tau_i - 1} - \left(\lambda \left(z_{t_i} + z_{t_{i-1}} r^{\tau_i} \right) - \frac{1}{u_{\tau_i}} \right) \frac{\partial u_{\tau_i}}{\partial r} + \left(\frac{\tau_i}{2r} + \frac{\frac{\partial u_{\tau_i}}{\partial r}}{u_{\tau_i}} \right) B(\theta : x_i) \right] \tag{5.10}$$

$$- \frac{1}{1-r} + \lambda z_0$$

and

$$\frac{\partial L(\theta)}{\partial \lambda}=\sum_{i=1}^{n}\left[-u_{\tau_i}\left(z_{t_i}+z_{t_{i-1}}r^{\tau_i}\right)+\frac{B(\theta:x_i)}{\lambda}\right]+\frac{(n+1)}{\lambda}-(1-r)z_0. \tag{5.11}$$

Then, $B_r(\theta:x_i)$, $B_\lambda(\theta:x_i)$ are the partial derivatives of $B(\theta:x_i)$ w.r.t. r and λ respectively. These are easily derived as;

$$B_r(\theta:x_i)=\left(\frac{\tau_i}{2r}+\frac{\frac{\partial u_{\tau_i}}{\partial r}}{u_{\tau_i}}\right)\left(B(\theta:x_i)+\frac{x_i^2}{I_0^2(x_i)}\left(I_0^2(x_i)-\frac{1}{x_i}I_0(x_i)I_1(x_i)-I_1^2(x_i)\right)\right) \tag{5.12}$$

and

$$B_\lambda(\theta:x_i)=\frac{1}{\lambda}\left(B(\theta:x_i)+\frac{x_i^2}{I_0^2(x_i)}\left(I_0^2(x_i)-\frac{1}{x_i}I_0(x_i)I_1(x_i)-I_1^2(x_i)\right)\right) \tag{5.13}$$

where

$$\frac{\partial u_{\tau_i}}{\partial r}=\frac{-u_{\tau_i}}{(1-r)}\left(1-u_{\tau_i}\tau_i r^{\tau_i-1}\right),\ \frac{\partial^2 u_{\tau_i}}{\partial r^2}=\frac{\tau_i r^{\tau_i-2}}{\left(1-r^{\tau_i}\right)}\left((\tau_i-1)u_{\tau_i}+2r\frac{\partial u_{\tau_i}}{\partial r}\right).$$

The second partial derivatives of L(θ) are

$$\frac{\partial^2 L(\theta)}{\partial r^2}=\sum_{i=1}^{n}\left[-\lambda u_{\tau_i}z_{t_{i-1}}\tau_i(\tau_i-1)r^{\tau_i-2}-\left(2\lambda z_{t_{i-1}}\tau_i r^{\tau_i-1}+\frac{\frac{\partial u_{\tau_i}}{\partial r}}{u_{\tau_i}^2}\right)\frac{\partial u_{\tau_i}}{\partial r}\right.$$

$$-\left(\lambda\left(z_{t_i}+z_{t_{i-1}}r^{\tau_i}\right)-\frac{1}{u_{\tau_i}}\right)\frac{\partial^2 u_{\tau_i}}{\partial r^2}-\left(\frac{\tau_i}{2r^2}+\frac{\left(\frac{\partial u_{\tau_i}}{\partial r}\right)^2}{u_{\tau_i}^2}-\frac{\frac{\partial^2 u_{\tau_i}}{\partial r^2}}{u_{\tau_i}}\right)B(\theta:x_i) \quad (5.14)$$

$$+\left(\frac{\tau_i}{2r}+\frac{\frac{\partial u_{\tau_i}}{\partial r}}{u_{\tau_i}}\right)B_r(\theta:x_i)\Bigg]-\frac{1}{(1-r)^2},$$

$$\frac{\partial^2 L(\theta)}{\partial\lambda dr}=\sum_{i=1}^{n}\left[-u_{\tau_i}z_{t_{i-1}}\tau_i r^{\tau_i-1}-\left(z_{t_i}+z_{t_{i-1}}r^{\tau_i}\right)\frac{\partial u_{\tau_i}}{\partial r}+\left(\frac{\tau_i}{2r}+\frac{\frac{\partial u_{\tau_i}}{\partial r}}{u_{\tau_i}}\right)B_\lambda(\theta:x_i)\right]+z_0, \quad (5.15)$$

$$\frac{\partial^2 L(\theta)}{\partial\lambda^2}=\sum_{i=1}^{n}\left[-\frac{B(\theta:x_i)}{\lambda^2}+\frac{B_\lambda(\theta:x_i)}{\lambda}\right]-\frac{(n+1)}{\lambda^2} \quad (5.16)$$

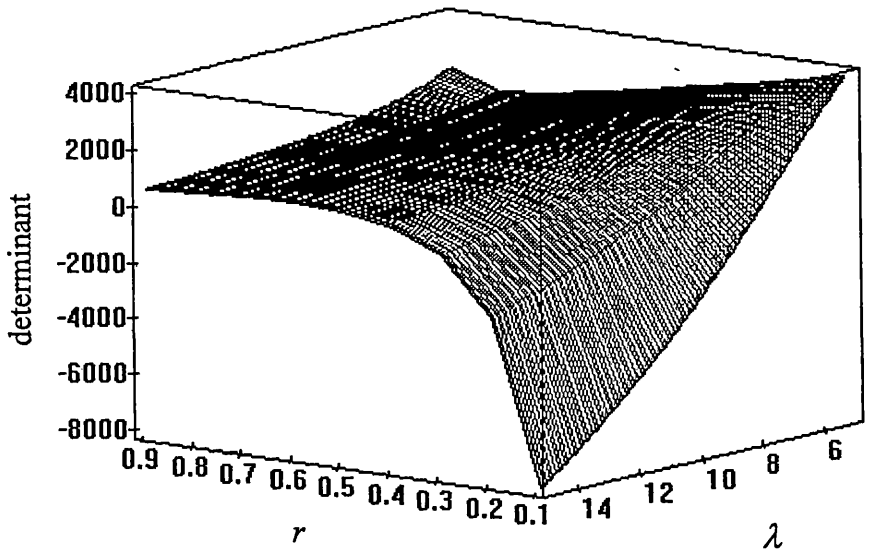
and

$$\frac{\partial^2 L(\theta)}{\partial r\partial\lambda}=\sum_{i=1}^{n}\left[-u_{\tau_i}z_{t_{i-1}}\tau_i r^{\tau_i-1}-\left(z_{t_i}+z_{t_{i-1}}r^{\tau_i}\right)\frac{\partial u_{\tau_i}}{\partial r}+\frac{B_r(\theta:x_i)}{\lambda}\right]+z_0. \quad (5.17)$$

To give some idea of the appearance of the determinant of the Hessian matrix, we used the Fortran 77 program as shown in Appendix C to simulate a time series of length t+1=500 observations from the Laguerre model with true values of (r, λ) = (0.5,10.0). Once again, 10% of these observations are deleted at random and substituted by zero. The determinant of the Hessian matrix, Δ, is then computed over the range

$r \in [0.1, 0.9]$ and $\lambda \in [5.0, 15.0]$. These results are shown in Figure (5.8) below. As discussed earlier, the first condition for a maximum point to exist is that $\Delta > 0$ at this point. However, $\Delta > 0$ in the vicinity of the true values of r and λ, with Δ only falling beneath zero for small values of r. Hence the first condition appears to hold.

Figure (5.8): Plot of Determinant of Hessian Matrix.



We can demonstrate, that the second derivative w.r.t. λ is negative over the range of r and λ. From (5.13), and (5.16).

$$\frac{\partial^2 L(\theta)}{\partial \lambda^2} = \sum_{i=1}^{n}\left[-\frac{B(\theta : x_i)}{\lambda^2} + \frac{B_\lambda(\theta : x_i)}{\lambda}\right] - \frac{(n+1)}{\lambda^2}$$

$$= \sum_{i=1}^{n}\left[\frac{x_i^2}{\lambda^2 I_0^2(x_i)}\left(I_0^2(x_i) - \frac{1}{x_i} I_0(x_i) I_1(x_i) - I_1^2(x_i)\right)\right] - \frac{(n+1)}{\lambda^2}.$$

Now, if we define $J(x) = \frac{1}{I_0^2(x)}\left[I_0^2(x) - \frac{1}{x} I_0(x) I_1(x) - I_1^2(x)\right]$ so that

$$\frac{\lambda^2 \partial^2 L(\theta)}{\partial \lambda^2} = \sum_{i=1}^{n} x_i^2 J(x_i) - (n+1) \; < \; n \max\left(x^2 J(x)\right) - (n+1) = nM - (n+1) \text{ say.}$$

So a necessary condition for $\frac{d^2L(\theta)}{d\lambda^2}$ to be negative is $M < \frac{n+1}{n}$. In the thesis by Beet (1995), M is found numerically to have a maximum value of approximately 0.67992 at $x \approx 2.4778$ and so M is certainly < 1 which is $< \frac{n+1}{n}$ for all $n \geq 1$. Therefore, $\frac{d^2L(\theta)}{d\lambda^2}$ would seem to be negative.

Chapter Six

Simulation and Analysis of Maximum Likelihood Estimates

In this chapter we investigate the statistical properties of the maximum likelihood estimators, $\hat{r}$ and $\hat{\lambda}$, for the Laguerre mixture model from time series with missing values. This is achieved through the medium of simulation. From large numbers of simulated series, we examine the sampling distributions of $\hat{r}$ and $\hat{\lambda}$ by using box-plots, histograms and normal Q-Q plots.

6.1 Simulation Run

To generate realisations of stationary sequences from the Laguerre model for given true values of r and λ, we use the following four steps.

(i) The first time series value z_0 is sampled from the stationary distribution, $\exp\{\lambda(1-r)\}$. This is achieved by using the NAG routine G05DBF.

(ii) A value i is chosen at random from the Poisson distribution with mean $r\lambda z_0$. This is achieved by using the NAG routines G05ECF and G05EYF.

(iii) The second time series value z_1 is sampled from a Gamma distribution with scale $1/\lambda$ and shape i+1. This is achieved by using the NAG routine G05DGF.

(iv) z_0 in part (ii) is replaced by the new time series value z_1 and the process is repeated.

This method is continued until the time series $z_0, z_1, ..., z_t$ is obtained. Then, to create a time series with missing values, we use the method explained in section 3.3 to substitute an exact proportion, known as δ, of the data observations $z_1, z_2, ..., z_{t-1}$ by zero's. Here, we introduce the notation r_0, λ_0 for the true values of the parameters r and λ.

In this preliminary investigation, we decided to examine 9 combinations for r_0 and λ_0, namely $r_0 \in (0,1)$, 0.05, 0.50 and 0.95 were employed with each λ_0 value of 1.0, 5.0 and 10.0. This is achieved by the Fortran 77 program shown in Appendix D, which generates for every combination of r_0 and λ_0, 100 realisations for each of the three lengths, 50, 250 and 1000 data observations. However, to perturb this series to a series which includes missing observations a value of δ=0.1 is used and hence, 10% of the data observations in the series are substituted at random for zero. Table (6.1) given overleaf show the results, in terms of summary statistics for every set of 100 realisations, the maximum likelihood estimates of r and λ i.e. $\hat{r}$ and $\hat{\lambda}$.

Table 6.1: Maximum likelihood simulation results[1] for a set of 50, 250 and 1000 data values.

r	λ	Mean of $\hat{r}$	Std dev of $\hat{r}$	Skewness of $\hat{r}$	Kurtosis of $\hat{r}$	Mean of $\hat{\lambda}$	Std dev of $\hat{\lambda}$	Skewness of $\hat{\lambda}$	Kurtosis of $\hat{\lambda}$	Corr($\hat{r},\hat{\lambda}$)
0.05	1.00	0.177	0.127	1.432	3.426	1.199	0.354	2.602	10.612	0.8079
	5.00	0.177	0.131	1.265	2.759	6.028	1.837	2.413	8.994	0.7988
	10.00	0.175	0.130	1.349	3.045	11.950	3.613	2.443	9.757	0.8173
0.50	1.00	0.495	0.139	-0.616	-0.164	1.061	0.269	0.366	-0.556	0.5261
	5.00	0.487	0.138	-0.578	-0.254	5.264	1.364	0.400	-0.580	0.5242
	10.00	0.491	0.138	-0.536	-0.200	10.588	2.710	0.362	-0.570	0.5130
0.95	1.00	0.932	0.050	-2.338	8.633	1.084	0.248	0.833	0.504	0.1350
	5.00	0.929	0.050	-1.972	6.856	5.399	1.298	1.017	1.240	0.1820
	10.00	0.931	0.049	-2.137	7.751	10.906	2.631	0.943	0.985	0.1797
0.05	1.00	0.086	0.053	0.861	0.942	1.045	0.095	1.233	4.345	0.6455
	5.00	0.084	0.053	0.912	1.102	5.215	0.471	1.355	4.635	0.6153
	10.00	0.084	0.054	0.821	0.906	10.438	0.972	1.113	3.946	0.6390
0.50	1.00	0.490	0.064	-0.414	0.494	1.004	0.104	0.357	-0.086	0.5626
	5.00	0.492	0.062	-0.266	0.156	5.055	0.531	0.210	-0.266	0.5883
	10.00	0.491	0.064	-0.422	0.558	10.091	1.026	0.298	-0.069	0.5641
0.95	1.00	0.944	0.020	-0.669	0.298	1.013	0.109	0.448	-0.069	0.1352
	5.00	0.943	0.021	-0.745	0.262	5.065	0.541	0.413	-0.124	0.2038
	10.00	0.944	0.021	-1.064	1.847	10.126	1.065	0.448	0.056	0.1520
0.05	1.00	0.054	0.033	0.134	-0.974	1.013	0.052	0.353	-0.329	0.7083
	5.00	0.053	0.033	0.181	-0.932	5.072	0.256	0.300	-0.313	0.7191
	10.00	0.053	0.032	0.120	-0.895	10.133	0.517	0.334	-0.364	0.7332
0.50	1.00	0.505	0.031	0.129	0.147	1.013	0.056	0.267	0.293	0.5395
	5.00	0.506	0.031	0.201	-0.011	5.062	0.275	0.326	0.432	0.5143
	10.00	0.505	0.031	0.263	0.106	10.132	0.554	0.286	0.357	0.5252
0.95	1.00	0.948	0.009	-0.259	-0.375	1.007	0.053	0.095	-0.219	0.1986
	5.00	0.948	0.008	-0.509	-0.145	5.029	0.277	0.093	-0.115	0.1883
	10.00	0.948	0.009	-0.348	-0.133	10.076	0.550	0.124	-0.077	0.1736

[1]S.E. Skewness for a sample of 100 = 0.241

S.E. Kurtosis for a sample of 100 = 0.478

Our results above indicate that as the length of the time series is increased by a multiple n, there is a factor reduction in standard deviation for the estimates $\hat{r}$ of approximately $1/\sqrt{n}$. However, this factor reduction in standard deviation seems to be initially faster for $\hat{\lambda}$ i.e. approximately $1/\sqrt{n^d}$ for $d>1$. The reduction in standard deviation for $\hat{r}$ is, in most cases by a factor of $1/\sqrt{5}$ as the series increases from 50 to 250 observations and approximately $1/\sqrt{4}$=0.5 as the series is increased from 250 to 10000.

Here, we define the moment ratio's

$$b_1=\frac{(m_3)^2}{(m_2)^3},\ b_2=\frac{m_4}{(m_2)^2}$$

where b_1 is the coefficient of skewness and b_2 is the coefficient of kurtosis. Both are expressed in terms of 2^{nd},3^{rd} and 4^{th} sample moments about the mean. These can be used to assess normality of the sample by measuring the extent to which the distributions of the estimates are asymmetric and tail heavy respectively. Rough tests of normality may be obtained by comparing b_1 and b_2 with the approximate vales of their standard errors[1], $\sqrt{(6/N)}$ and $\sqrt{(24/N)}$ respectively.

The tabulated results show the distribution of $\hat{r}$ for small sample sizes to be skewed in the positive direction for r_0=0.05 but for a value of r_0=0.95, the estimates are negatively skewed. Hence, as the value of r_0 shifts across 0.5 a symmetry seems to emerge with the direction of skewness changing from positive to negative. A possible explanation for the skewness of $\hat{r}$ could be that it possesses two boundaries and with r_0 values near to either boundary, skewness is inevitable. However, as the length of the series increases to 1000 observations, the coefficient of skewness values become less significant[1] with a few having skewness values extremely close to zero indicating near-perfect symmetrical distributions. In addition, a pattern seems to emerge when examining the type of kurtosis for the distribution of $\hat{r}$, since for small sample sizes $\hat{r}$ has a leptokurtic distribution i.e. $b_2>0$ but as t increases the distribution changes to

platykurtic i.e. $b_2 < 0$. However, both the coefficients of skewness and kurtosis of $\hat{r}$ and $\hat{\lambda}$ become less significant[1] with increasing sample size which is analogous for any value of r_0 and λ_0. Hence, as the sample size increases, the estimators, $\hat{r}$ and $\hat{\lambda}$ approach their true values, in probability and become consistent. For example, if the assumption that $\hat{r}$ is a consistent estimator for r_0 holds, then for any positive constant c

$$\lim_{t \to \infty} \mathrm{P}\left(\left|\hat{r} - r_0\right| \geq c\right) = 0$$

where t denotes the size of the sample.

The suggestion that $\hat{r}$ is a consistent estimator is supported by the boxplots give in Figure (6.1) below, which show the distribution of $\hat{r}$ for specified values of r_0, λ_0 over varying sample sizes. The variance reduction associated with increasing sample sizes shown in Table (6.1) is now visually demonstrated. This has an effect on the inter-quartile range, which can be seen to decrease as the length of the time series increases to 1000 observations. Examination also shows, that a value of r_0 at which a symmetry appears to emerge with the direction of skewness changing from positive to negative appears to be slightly lower at ≈ 0.4. This is so, since for small sample sizes the distribution of $\hat{r}$ continues to be skewed in the negative direction for a value of r_0=0.5. However, as the sample size increases the coefficient of skewness becomes less significant[1] with the distribution of $\hat{r}$ approaching a normal distribution.

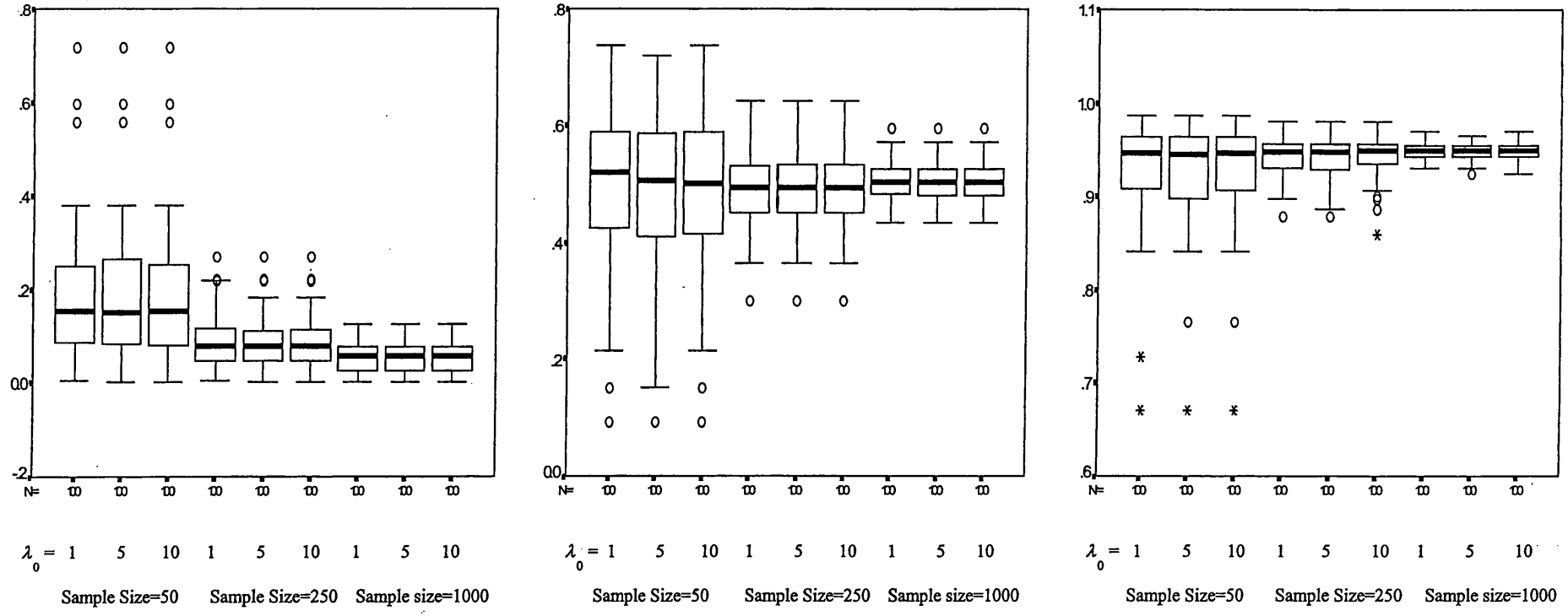
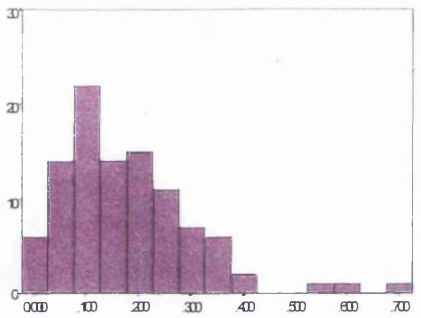
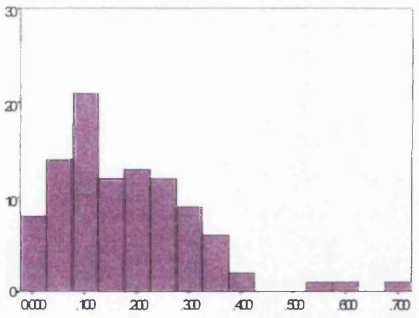


Figure (6.1): Boxplots of Maximum Likelihood Estimator $\hat{r}$, for r_0=0.05, 0.50, 0.95 and λ_0=1.0, 5.0, 10.0 for Series of Length 50, 250 and 1000.

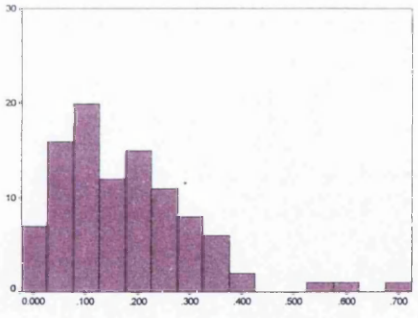
Figures (6.2) to (6.7) respectively show frequency histograms and normal Q-Q plots of $\hat{r}$ for specified values of r_0, λ_0 and time series of varying lengths. The distribution of $\hat{r}$ for r_0=0.05 is positively skewed i.e. $\mathrm{E}(\hat{r}) - r_0 > 0$ and unimodal with a longer tail towards the increasing x-axis but negatively skewed for r_0=0.95. Also, a pattern seems to emerge when examining the type of kurtosis in the $\hat{r}$ distribution since for small sample sizes $\hat{r}$ possess a leptokurtic distribution characterised by a sharper peak at the mean and more extended tails. In contrast, as t increases $\hat{r}$ changes to a platykurtic distribution producing a flatter top and more abrupt terminals, However, for larger sample sizes the coefficient of skewness and kurtosis associated with each histogram become less significant[1] and this is analogous for the distribution of $\hat{r}$ given any r_0, λ_0 value. Hence, if a large enough set of data observations are used it seems that the distribution of $\hat{r}$ approaches a normal distribution. This approach to normality can be expected to be much slower for values of r_0 close to zero because of the influence of the boundary. The results illustrated by the frequency histograms are also reflected in the normal Q-Q plots.



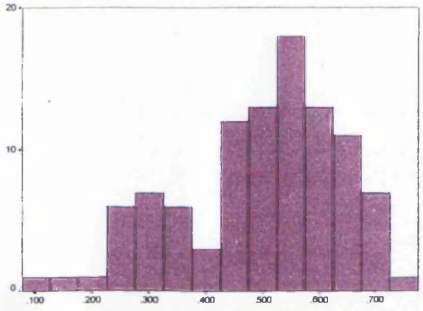
r_0=0.05 λ_0=1.0



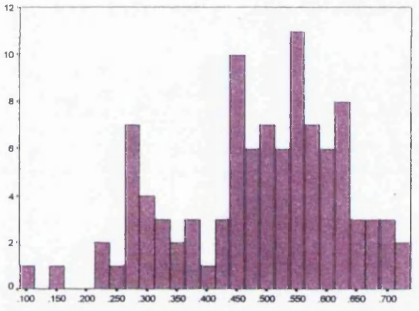
r_0=0.05 λ_0=5.0



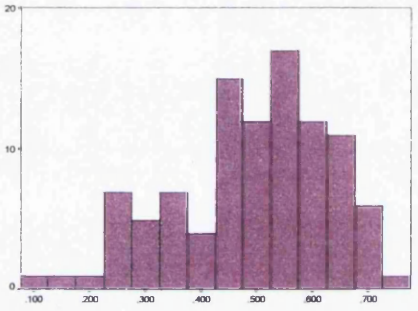
r_0=0.05 λ_0=10.0



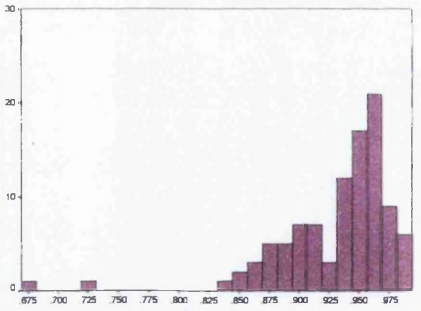
r_0=0.50 λ_0=1.0



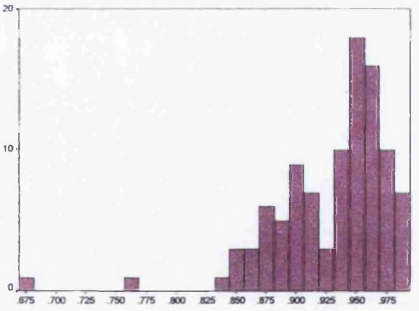
r_0=0.50 λ_0=5.0



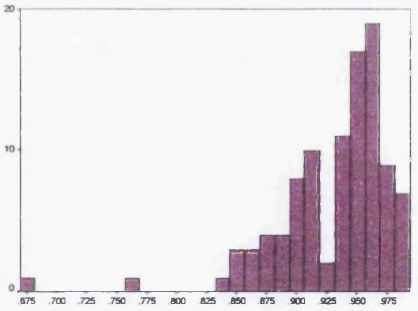
r_0=0.50 λ_0=10.0



r_0=0.95 λ_0=1.0

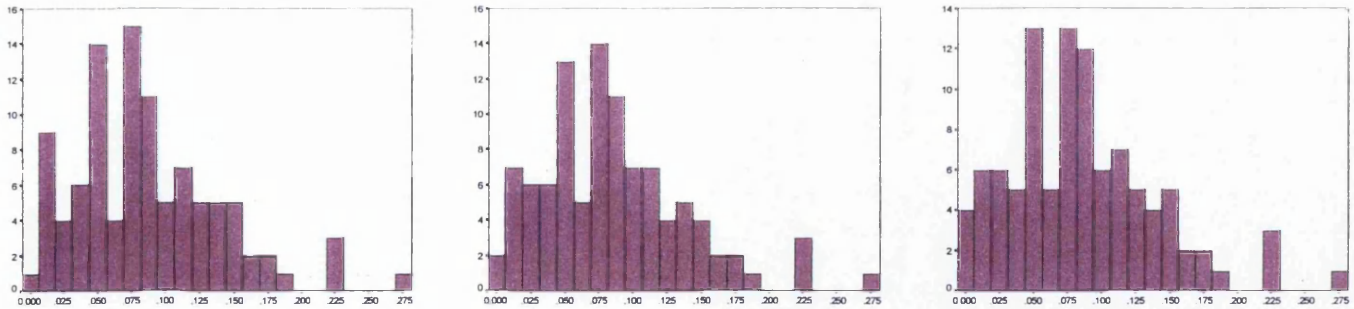


r_0=0.95 λ_0=5.0

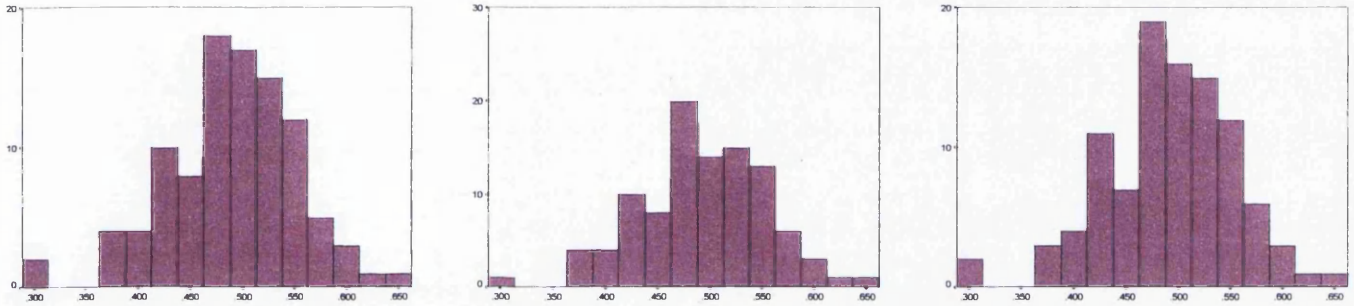


r_0=0.95 λ_0=10.0

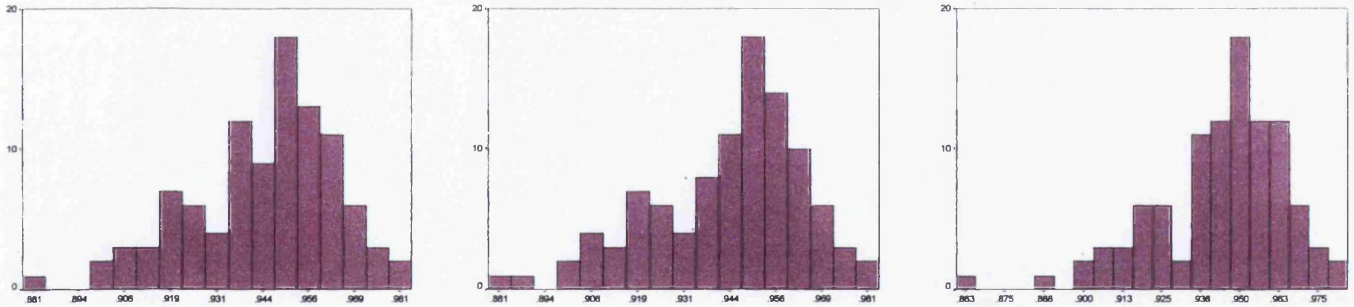
Figure (6.2): Frequency Histogram of Maximum Likelihood Estimator $\hat{r}$, for r_0=0.05,0.50,0.95 and λ_0=1.0,5.0,10.0 for a series of length 50.



r_0=0.05 λ_0=1.0 r_0=0.05 λ_0=5.0 r_0=0.05 λ_0=10.0

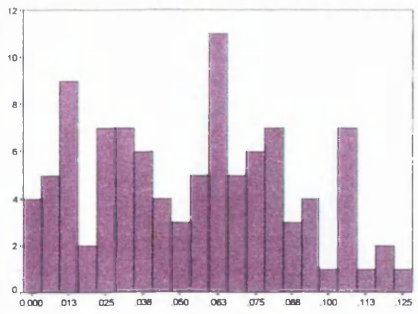


r_0=0.50 λ_0=1.0 r_0=0.50 λ_0=5.0 r_0=0.50 λ_0=10.0

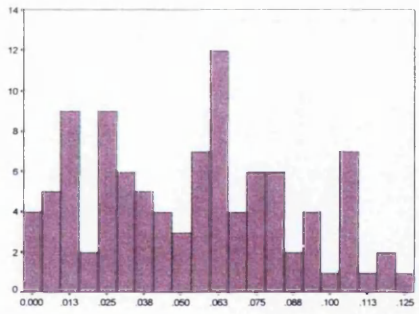


r_0=0.95 λ_0=1.0 r_0=0.95 λ_0=5.0 r_0=0.95 λ_0=10.0

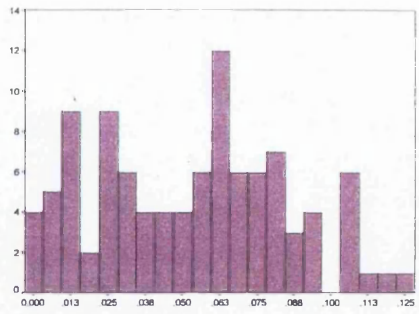
Figure (6.3): Frequency Histogram of Maximum Likelihood Estimator $\hat{r}$, for r_0=0.05,0.50,0.95 and λ_0=1.0,5.0,10.0 for a series of length 250.



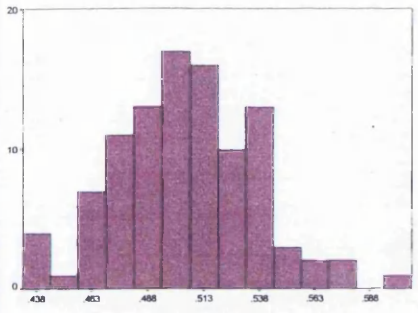
r_0=0.05 λ_0=1.0



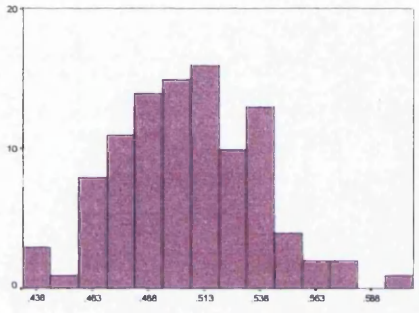
r_0=0.05 λ_0=5.0



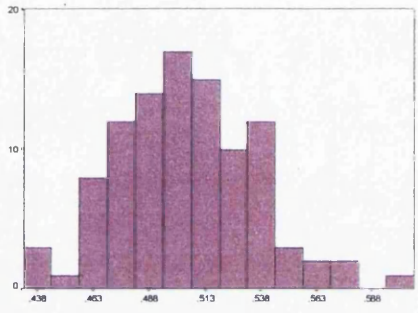
r_0=0.05 λ_0=10.0



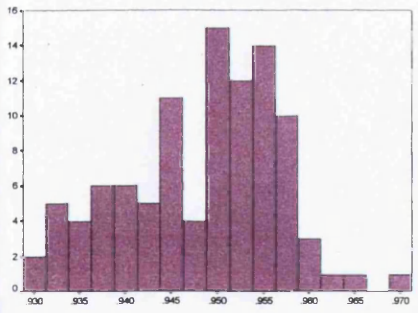
r_0=0.50 λ_0=1.0



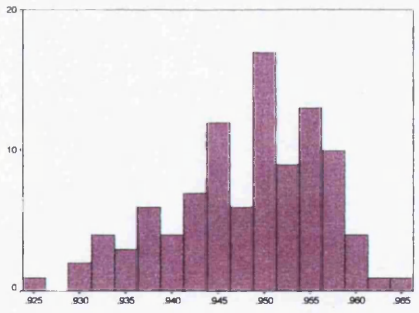
r_0=0.50 λ_0=5.0



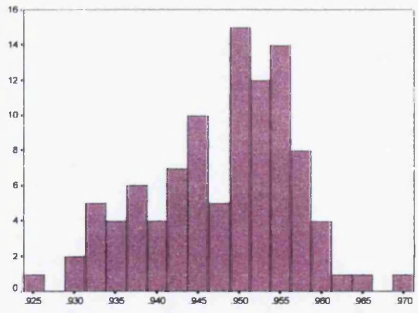
r_0=0.50 λ_0=10.0



r_0=0.95 λ_0=1.0

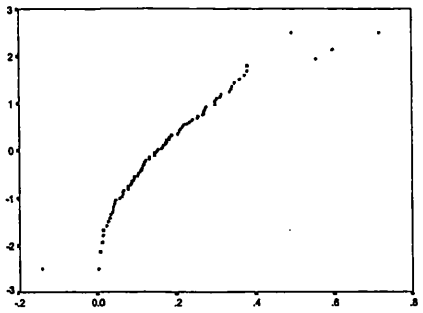


r_0=0.95 λ_0=5.0

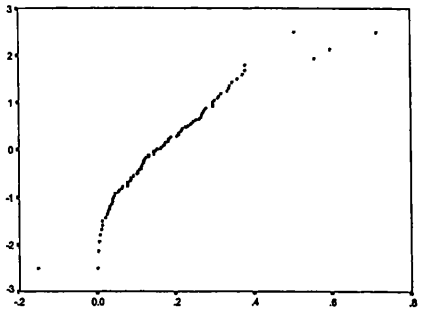


r_0=0.95 λ_0=10.0

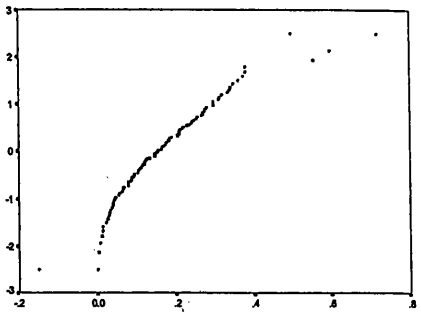
Figure (6.4): Frequency Histogram of Maximum Likelihood Estimator $\hat{r}$, for r_0=0.05,0.50,0.95 and λ_0=1.0,5.0,10.0 for a series of length 1000.



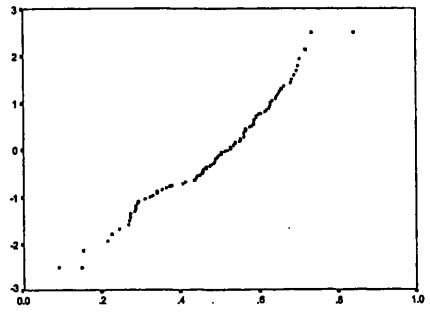
r_0=0.05 λ_0=1.0



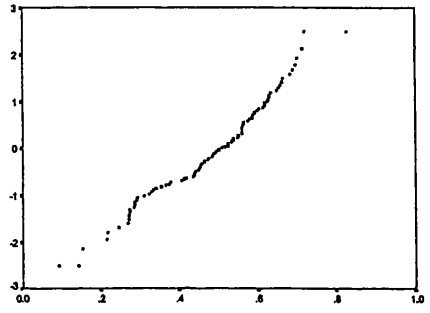
r_0=0.05 λ_0=5.0



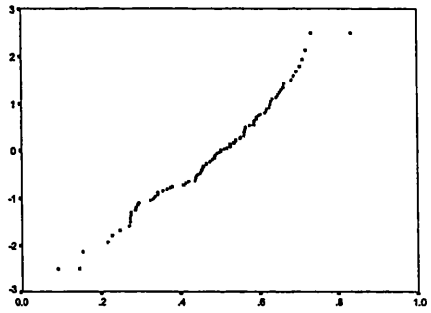
r_0=0.05 λ_0=10.0



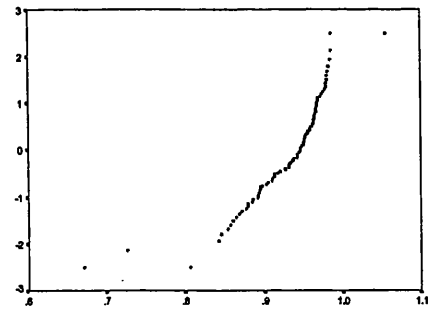
r_0=0.50 λ_0=1.0



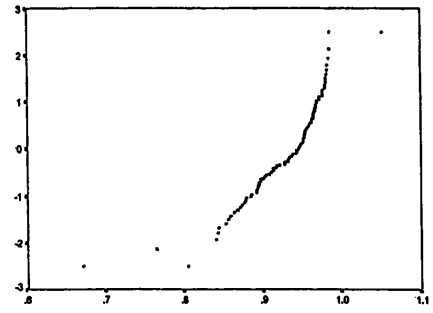
r_0=0.50 λ_0=5.0



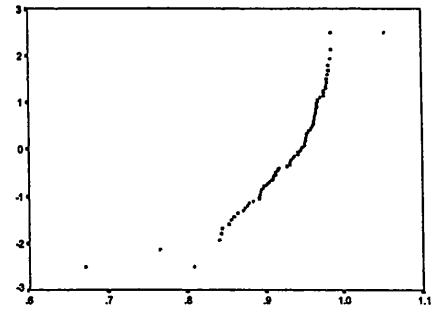
r_0=0.50 λ_0=10.0



r_0=0.95 λ_0=1.0

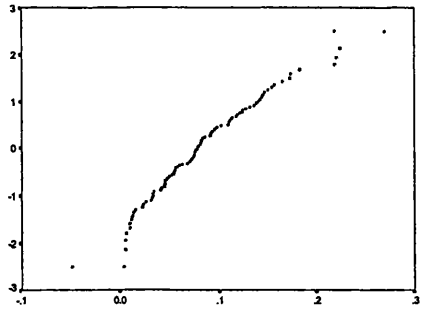


r_0=0.95 λ_0=5.0

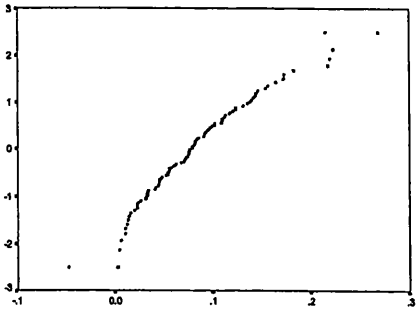


r_0=0.95 λ_0=10.0

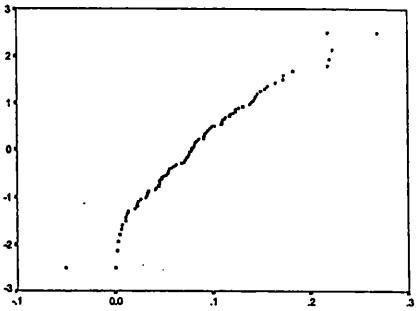
Figure (6.5): Normal Q-Q Plots of Maximum Likelihood Estimator $\hat{r}$, for r_0=0.05, 0.50, 0.95 and λ_0=1.0, 5.0, 10.0 for a series of length 50.



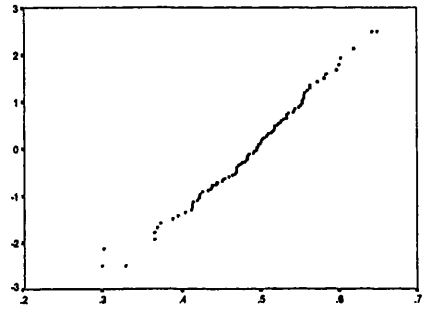
r_0=0.05 λ_0=1.0



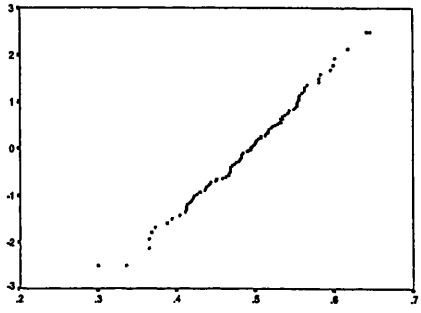
r_0=0.05 λ_0=5.0



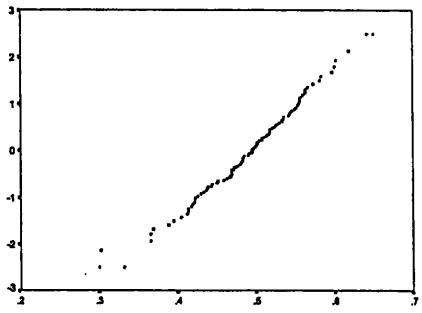
r_0=0.05 λ_0=10.0



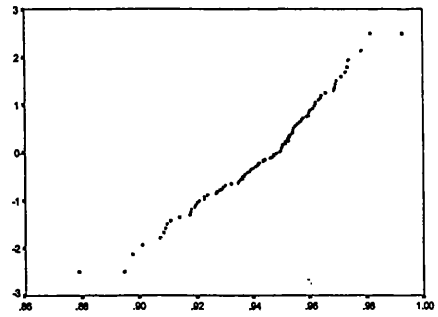
r_0=0.50 λ_0=1.0



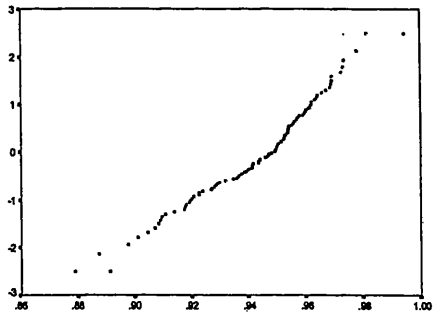
r_0=0.50 λ_0=5.0



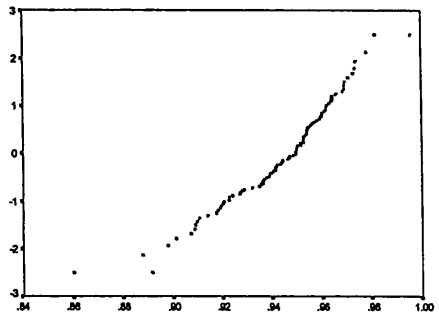
r_0=0.50 λ_0=10.0



r_0=0.95 λ_0=1.0

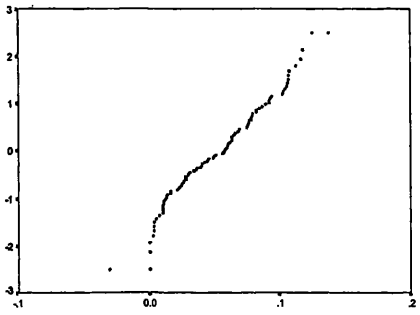


r_0=0.95 λ_0=5.0

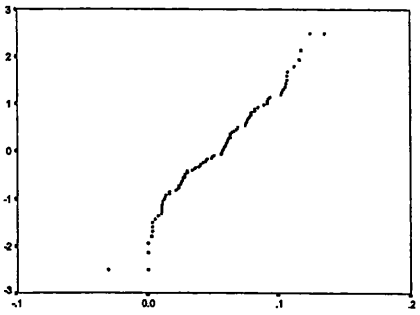


r_0=0.95 λ_0=10.0

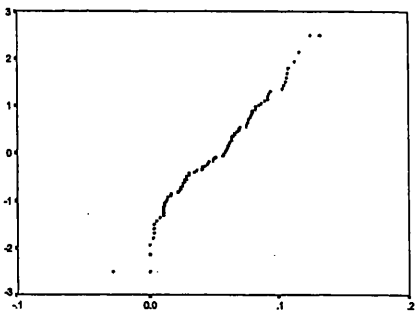
Figure (6.6): Normal Q-Q Plots of Maximum Likelihood Estimator $\hat{r}$, for r_0=0.05, 0.50, 0.95 and λ_0=1.0, 5.0, 10.0 for a series of length 250.



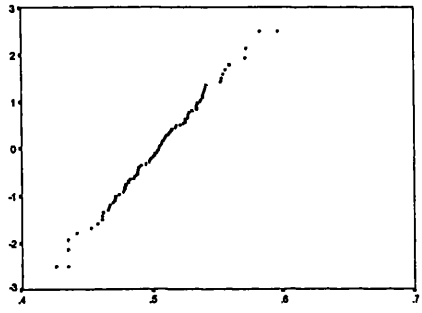
r_0=0.05 λ_0=1.0



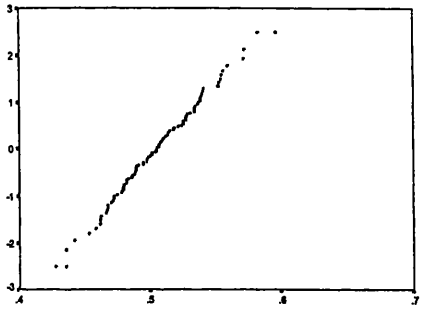
r_0=0.05 λ_0=5.0



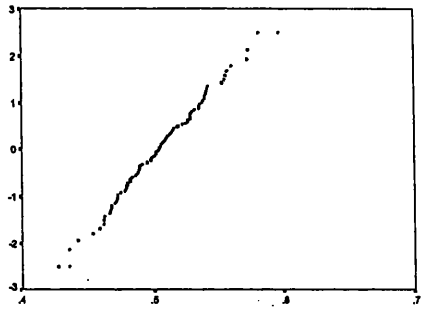
r_0=0.05 λ_0=10.0



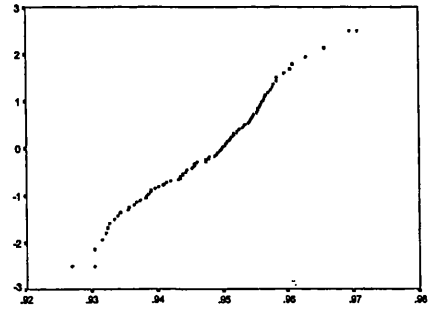
r_0=0.50 λ_0=1.0



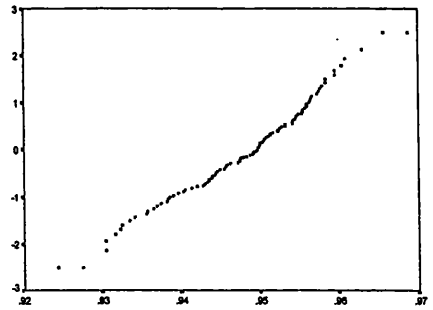
r_0=0.50 λ_0=5.0



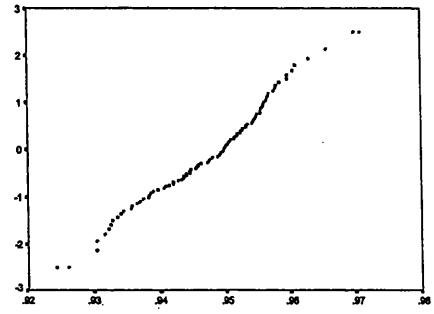
r_0=0.50 λ_0=10.0



r_0=0.95 λ_0=1.0



r_0=0.95 λ_0=5.0



r_0=0.95 λ_0=10.0

Figure (6.7): Normal Q-Q Plots of Maximum Likelihood Estimator $\hat{r}$, for r_0=0.05, 0.50, 0.95 and λ_0=1.0, 5.0, 10.0 for a series of length 1000.

The boxplots given in Figure (6.8) overleaf illustrate the improving accuracy of the λ estimates as the sample size increases. Again, as the length of the series is increased, the variance of λ estimates decreases and the coefficient of skewness and kutosis values associated with each histogram become less significant[1], with a tendency towards the normal distribution appearing. However, this seems to be far quicker for the estimators $\hat{\lambda}$ as opposed for $\hat{r}$. This has been observed for the distribution of $\hat{\lambda}$ for any value of r_0, λ_0.

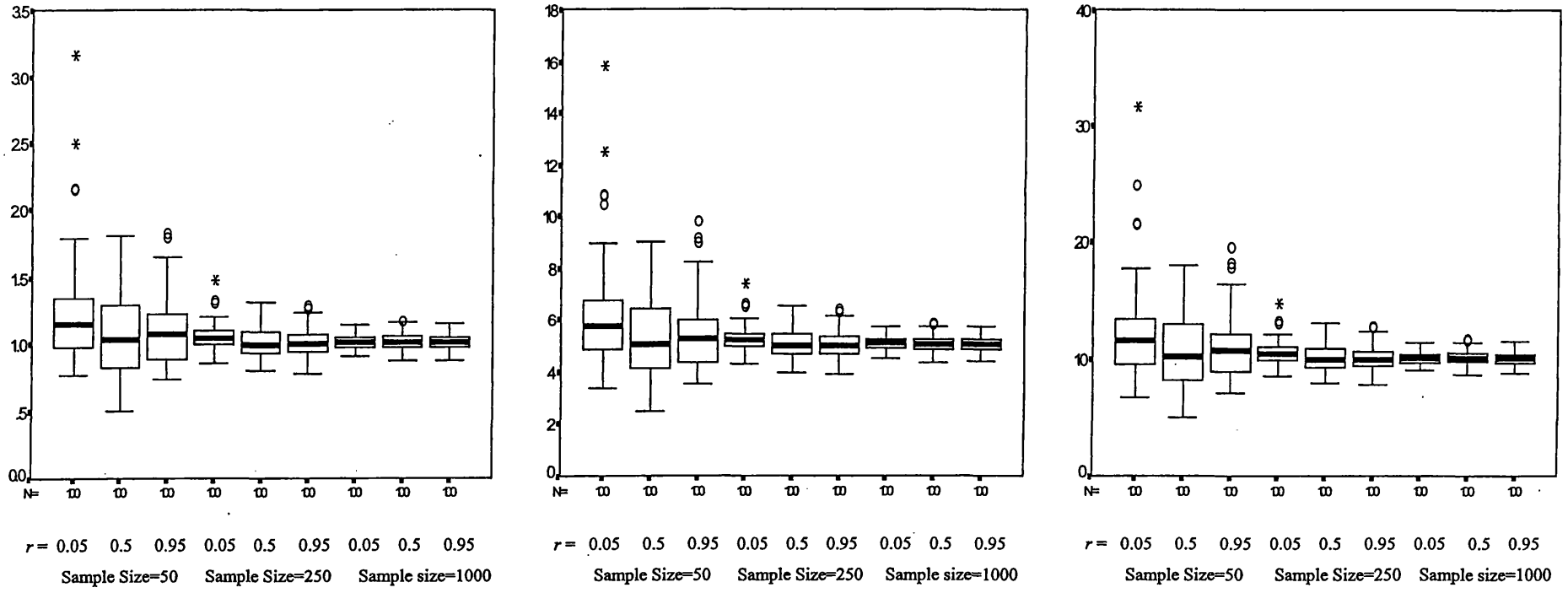
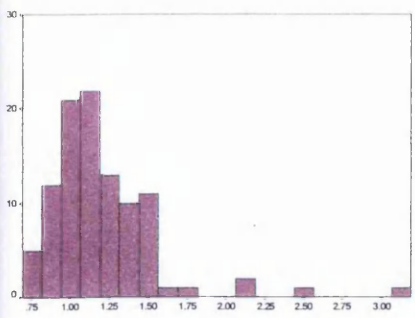
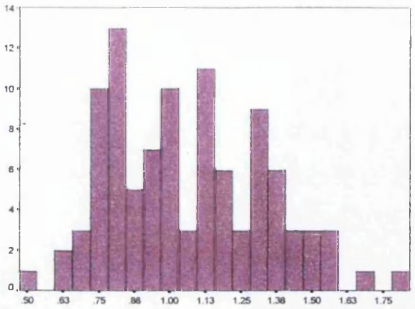


Figure (6.8): Boxplots of maximum likelihood estimator $\hat{\lambda}$, for λ_0=1.0, 5.0, 10.0 and r_0=0.05, 0.50, 0.95 for series of length 50, 250 and 1000.

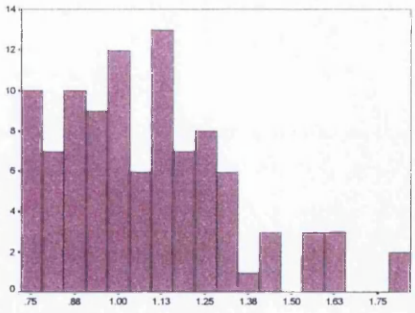
Figures (6.9) to (6.14) respectively show frequency histograms and normal Q-Q plots of $\hat{\lambda}$ for specified values of r_0, λ_0 over varying sample sizes. The frequency histograms illustrate the distribution of $\hat{\lambda}$ to be strongly positive skewed for short time series but as the length of the series increases to 1000, the coefficient of skewness become less significant[1] with a few distributions having vales extremely close to zero indicating near perfect symmetrical distributions. In addition, the coefficient of kurtosis becomes less significant[1], as the length of the time series increases this is analogous for the distribution of $\hat{\lambda}$ given any r_0, λ_0. However, for values of r_0=0.05, 0.95 and any value of λ_0, $\hat{\lambda}$ has a leptokurtic distribution for small sample sizes but as the sample size increases the distribution changes to platykurtic. Although, the opposite pattern seems to emerge for a value of r_0=0.5 with the type of the kurtosis changing from a platykurtic distribution to leptokurtic. These results are also reflected in the normal Q-Q plots.



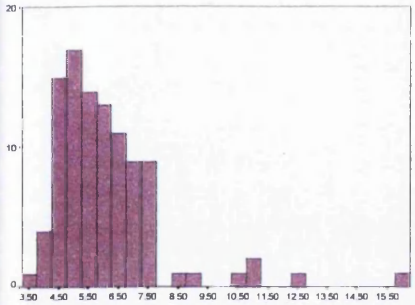
λ_0=1.0 r_0=0.05



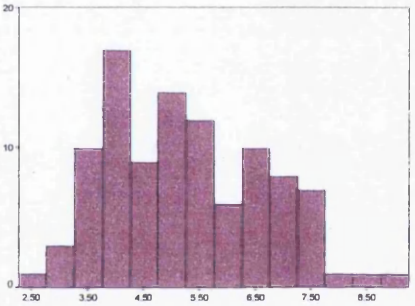
λ_0=1.0 r_0=0.50



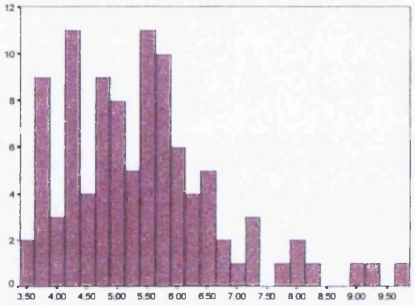
λ_0=1.0 r_0=0.95



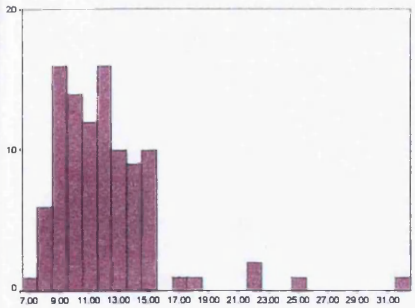
λ_0=5.0 r_0=0.05



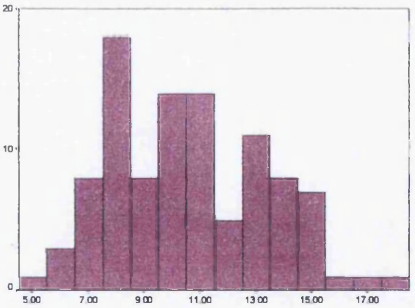
λ_0=5.0 r_0=0.50



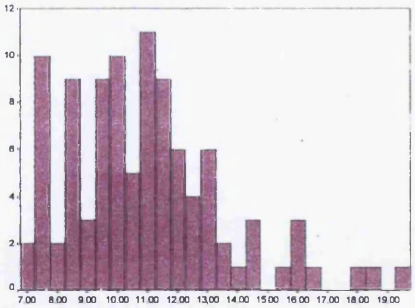
λ_0=5.0 r_0=0.95



λ_0=10.0 r_0=0.05

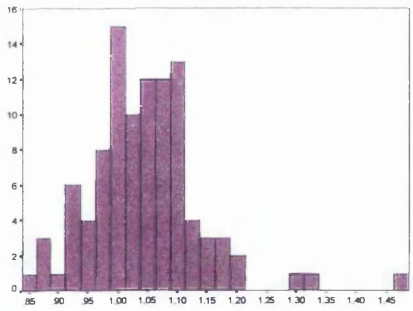


λ_0=10.0 r_0=0.50

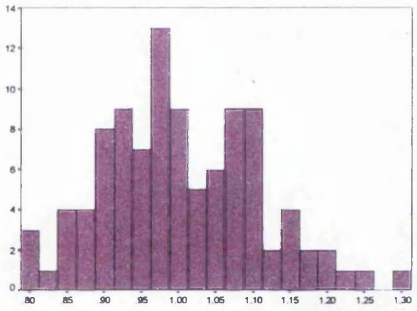


λ_0=10.0 r_0=0.95

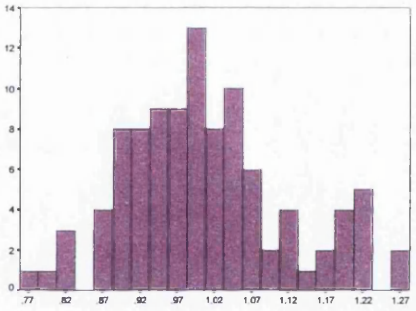
Figure (6.9): Frequency Histogram of Maximum Likelihood Estimator $\hat{\lambda}$, for λ_0=1.0, 5.0, 10.0 and r_0=0.05, 0.50, 0.95 for a series of length 50.



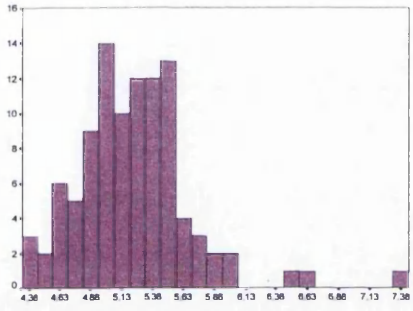
λ_0=1.0 r_0=0.05



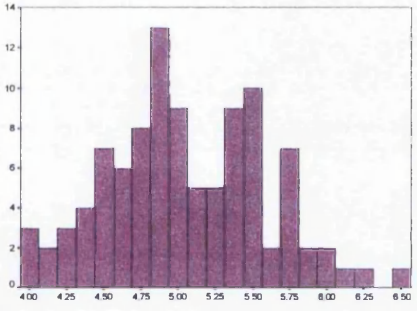
λ_0=1.0 r_0=0.50



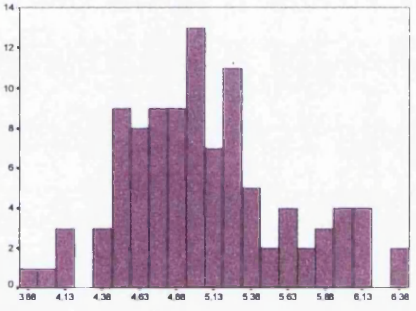
λ_0=1.0 r_0=0.95



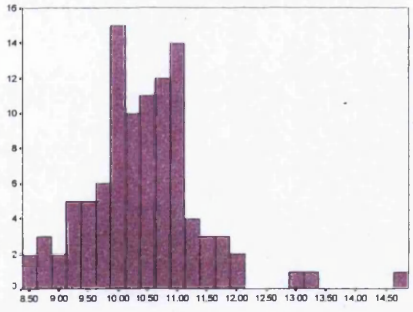
λ_0=5.0 r_0=0.05



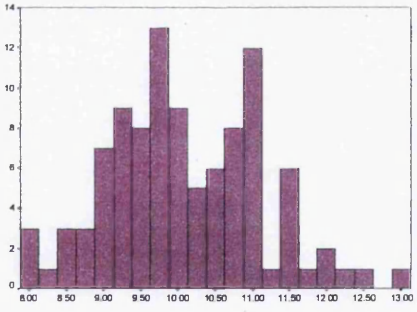
λ_0=5.0 r_0=0.50



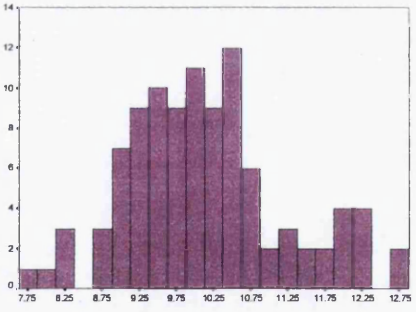
λ_0=5.0 r_0=0.95



λ_0=10.0 r_0=0.05

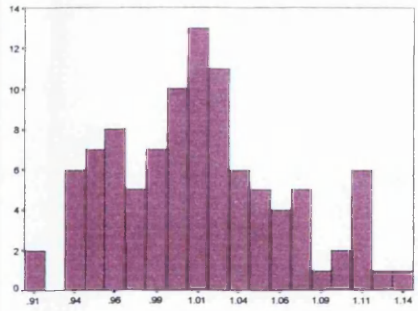


λ_0=10.0 r_0=0.50

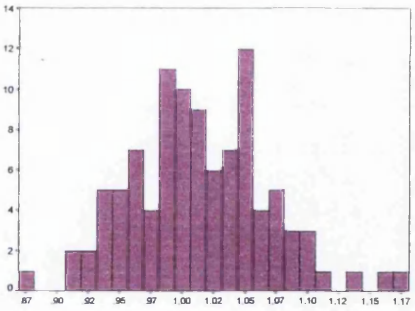


λ_0=10.0 r_0=0.95

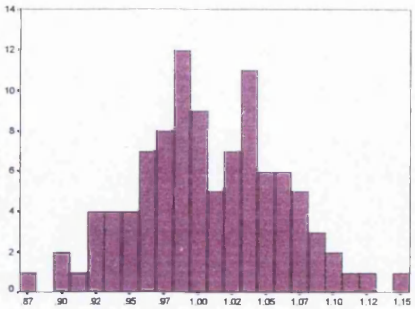
Figure (6.10): Frequency Histogram of Maximum Likelihood Estimator $\hat{\lambda}$, for λ_0=1.0, 5.0, 10.0 and r_0=0.05, 0.50, 0.95 for a series of length 250.



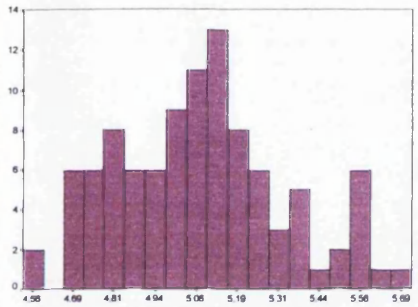
λ_0=1.0 r_0=0.05



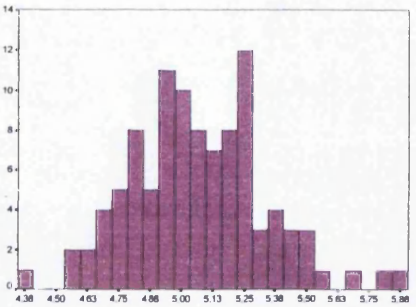
λ_0=1.0 r_0=0.50



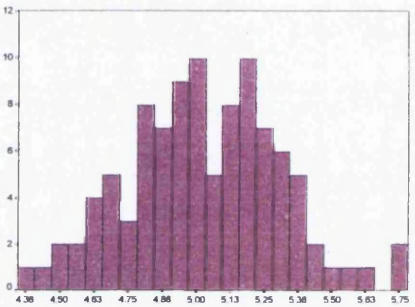
λ_0=1.0 r_0=0.95



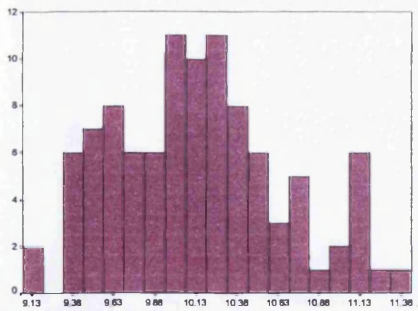
λ_0=5.0 r_0=0.05



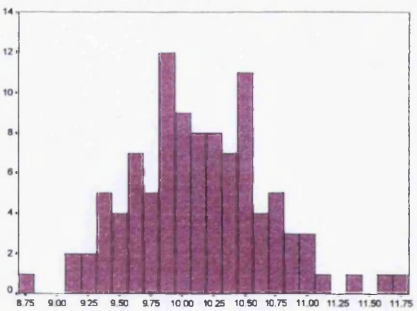
λ_0=5.0 r_0=0.50



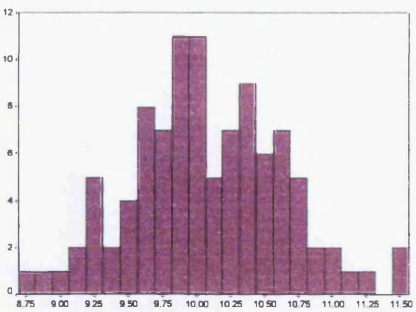
λ_0=5.0 r_0=0.95



λ_0=10.0 r_0=0.05

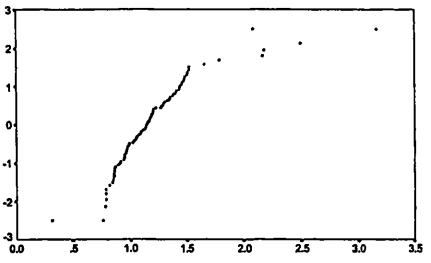


λ_0=10.0 r_0=0.50

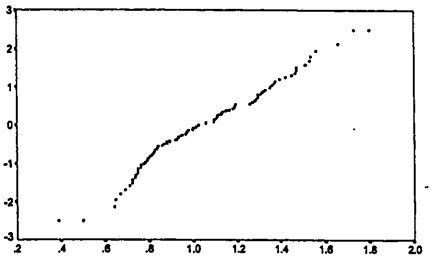


λ_0=10.0 r_0=0.95

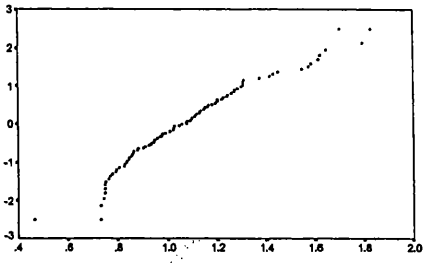
Figure (6.11): Frequency Histogram of Maximum Likelihood Estimator $\hat{\lambda}$, for λ_0=1.0, 5.0, 10.0 and r_0=0.05, 0.50, 0.95 for a series of length 1000.



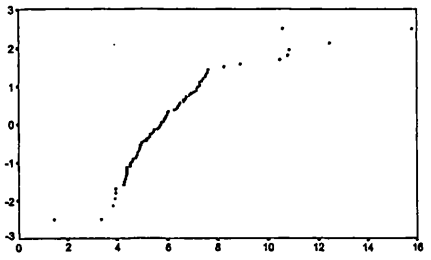
λ_0=1.0 r_0=0.05



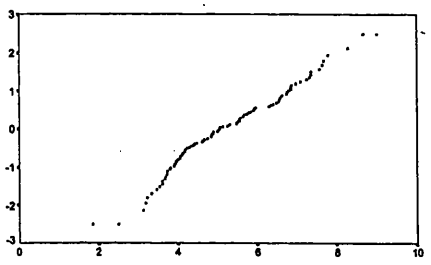
λ_0=1.0 r_0=0.50



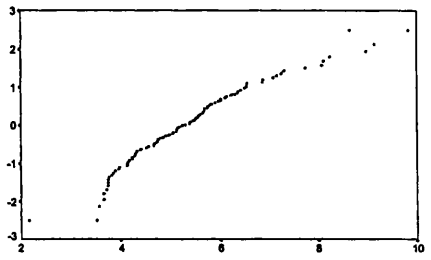
λ_0=1.0 r_0=0.95



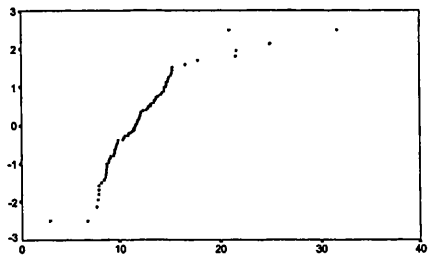
λ_0=5.0 r_0=0.05



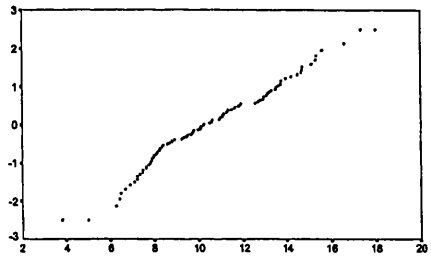
λ_0=5.0 r_0=0.50



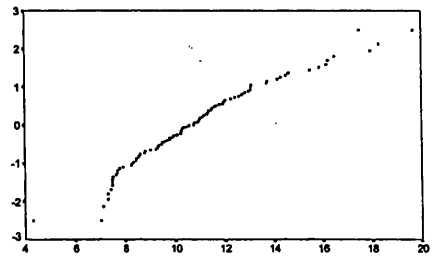
λ_0=5.0 r_0=0.95



λ_0=10.0 r_0=0.05

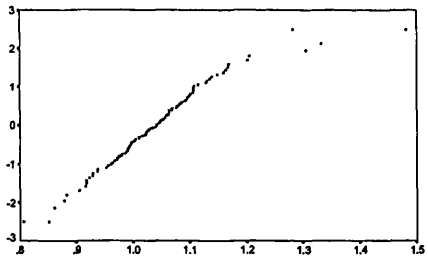


λ_0=10.0 r_0=0.50

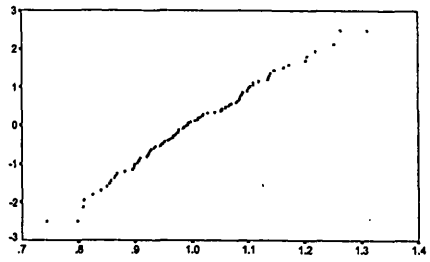


λ_0=10.0 r_0=0.95

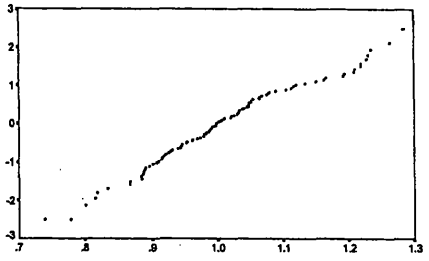
Figure (6.12): Normal Q-Q Plots of Maximum Likelihood Estimator $\hat{\lambda}$, for λ_0=1.0, 5.0, 10.0 and r_0=0.05, 0.50, 0.95 for a series of length 50.



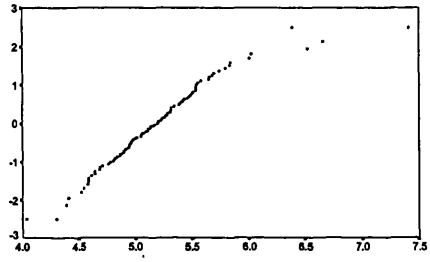
λ_0=1.0 r_0=0.05



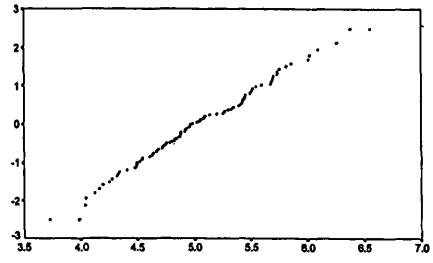
λ_0=1.0 r_0=0.50



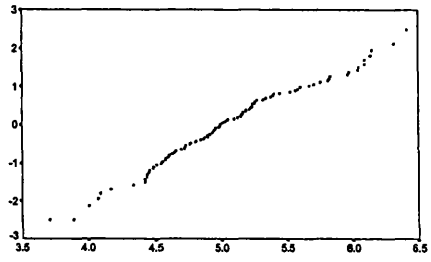
λ_0=1.0 r_0=0.95



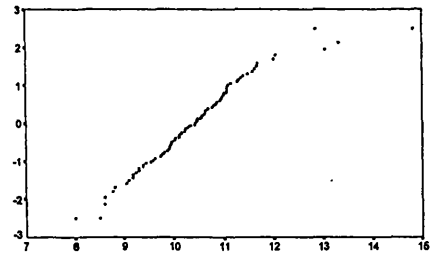
λ_0=5.0 r_0=0.05



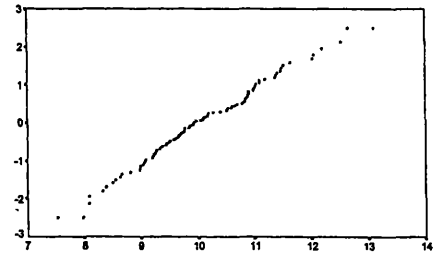
λ_0=5.0 r_0=0.50



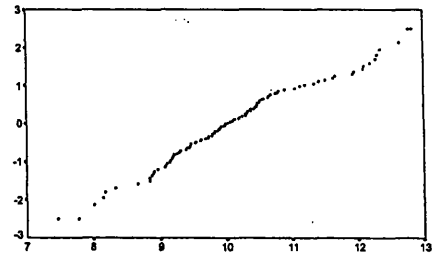
λ_0=5.0 r_0=0.95



λ_0=10.0 r_0=0.05

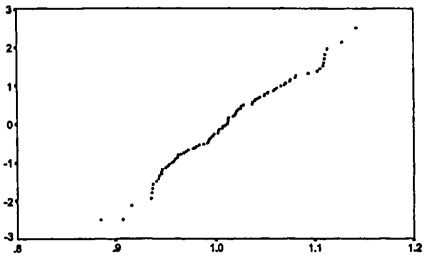


λ_0=10.0 r_0=0.50

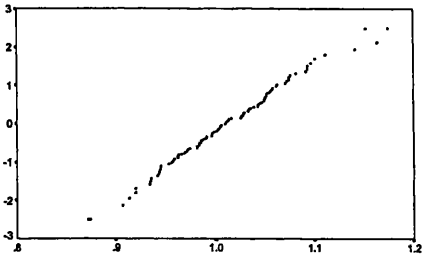


λ_0=10.0 r_0=0.95

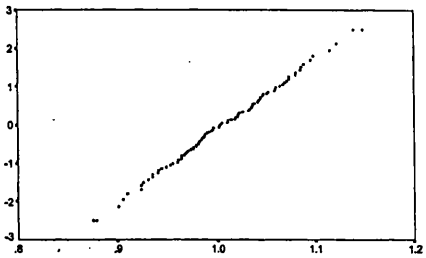
Figure (6.13): Normal Q-Q Plots of Maximum Likelihood Estimator $\hat{\lambda}$, for λ_0=1.0, 5.0, 10.0 and r_0=0.05, 0.50, 0.95 for a series of length 250.



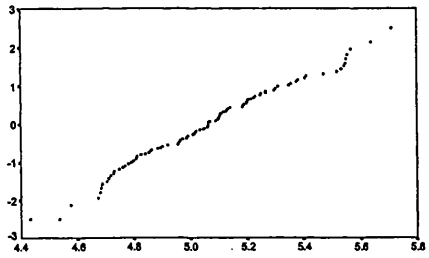
λ_0=1.0 r_0=0.05



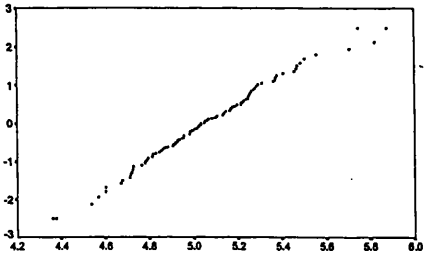
λ_0=1.0 r_0=0.50



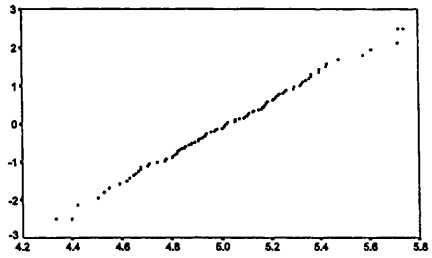
λ_0=1.0 r_0=0.95



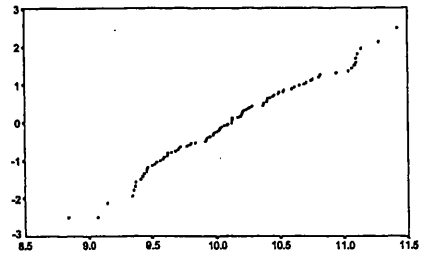
λ_0=5.0 r_0=0.05



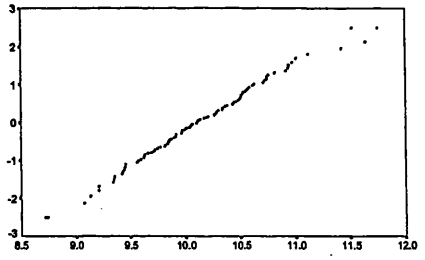
λ_0=5.0 r_0=0.50



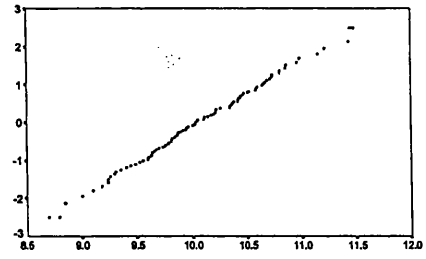
λ_0=5.0 r_0=0.95



λ_0=10.0 r_0=0.05



λ_0=10.0 r_0=0.50



λ_0=10.0 r_0=0.95

Figure (6.14): Normal Q-Q Plots of Maximum Likelihood Estimator $\hat{\lambda}$, for λ_0=1.0, 5.0, 10.0 and r_0=0.05, 0.50, 0.95 for a series of length 1000.

Chapter Seven

Conclusions

In recent years, we have witnessed an ever-increasing interest in the subject of time series analysis. In an attempt to provide a more real approximation to the world around us, analysts have moved away from linear time series and hence the linear constraint imposed on models to develop a new class of non-linear modelling.

In this thesis, we examine a new class of non-linear models known as mixture models, as developed by Jalali and Pemberton (1995). In particular, we focused on the Laguerre mixture model. We investigated the likelihood function of the Laguerre mixture model for a time series with missing values and found that due to the complexity of the first

derivatives of $L(\theta)$, we were unable to find a critical point analytically. Therefore we investigated whether a maximum point exists through examining plots and numerical optimisation routines. Our results indicated that a unique maximum point does exist.

Finally, we examine the statistical properties of the maximum likelihood estimators $\hat{r}$ and $\hat{\lambda}$, for the Laguerre mixture model from a time series with missing values. In particular we examined box-plots, histograms and Q-Q plots of both $\hat{r}$ and $\hat{\lambda}$ for specified values of r_0 and λ_0 with time series of varied lengths.

We found that the distribution of $\hat{r}$ for small sample sizes to be skewed in the positive direction for $r_0 = 0.05$ and negatively skewed for $r_0 = 0.95$. We interpreted this observed skewness of $\hat{r}$ as being a consequence of possessing two boundaries, and with r_0 near to either boundary, skewness is inevitable. In addition, we found the distribution of $\hat{r}$ for small sample sizes to be leptokurtic but as t increases the distribution changes to platykurtic. However as the sample size increases both coefficients of skewness and kurtosis of $\hat{r}$ became less significant, with the distribution of $\hat{r}$ approaching a normal distribution. However, due to the influence of the boundary this approach to normality can be expected to be much slower for values of r_0 close to zero.

Further, we found the distribution of $\hat{\lambda}$ to be strongly positive skewed for short time series but as the length of the time series increases the coefficient of skewness became less significant, with a few distributions indicating near perfect symmetry. In addition, as the length of the time series increased the coefficient of kurtosis became less significant. However, for values of $r_0 = 0.05, 0.95$ and any value of λ_0, the distribution of $\hat{\lambda}$ is leptokurtic for small sample sizes, but as the sample size increases the distribution changes to platykurtic. In contrast, for a value of $r_0 = 0.5$, the distribution changes from platykurtic to leptokurtic. Finally, although as the length of the time series increases the distributions of both $\hat{r}$ and $\hat{\lambda}$ approach the normal distribution, we found this to be quicker for the estimators $\hat{\lambda}$ as opposed to $\hat{r}$.

Bibliography

Anderson, T.W. (1971), The Statistical Analysis of Time Series, John Wiley and Sons inc, New York.

Beet, A.R. (1995), Mixture Models for Non-Linear Time Series. Thesis (M.Phil), University of Wales, Swansea.

Brillinger, D.R. (1975), Time Series: Data Analysis and Theory, Holt, Rinehart and Winston, New York.

Brockett, R.W. (1976), Volterra Series and Geometric Control Theory, Automatica, 12: 167-76.

Fisher, Sir R.A. (1922), Statistical Methods for Research Workers. Oliver and Boyd, London.

Franses, P.H. and Dijk, D.V. (2000), Non-Linear Time Series Models in Empirical Finance, Cambridge University Press, ISBN: 0521779650.

Granger, C.W.J. and Anderson, A.P. (1978), An Introduction to Bilinear Time Series Models. Vandenhoeck and Ruprecht, Gottingen and Zurich.

Jalali, A. and Pemberton, J. (1995), Mixture Models for Time Series. J.Appl.Prob., 32:123-38.

Jones, R.H. (1980), Maxium Likelihood Fitting of ARMA Models to Time Series with Missing Observations. Technometrics ., 22: 389-95.

Hagan, V. and Ozaki, T. (1981), Modelling Non-Linear Random Vibrations Using an Amplitude Dependent Autoregression Time Series Model. Biometrika., 68:189-96.

Hannan, E.J. (1970), Multilpe Time Series, John Wiley, New York.

Hannan, E.J. and Heyde, C.C. (1972), On Limit Theorems for Quadratic Functions of Discrete Time Series. Ann.Math. Statist,., 43: 2058-66.

Hannan, E.J. (1986), Rememberance of Things Past. The Craft of Probabilistic Modelling, (ed. J.Gani), Springer Verlag:31-42.

Hutton, J.L. (1990), Non-Negative Time Series Models for Dry River Flow, J. Appl.Prob., 27: 171-82.

Kantz, H. and Schreiber, T. (2003), Nonlinear Time Series Analysis, Cambridge University Press, ISBN:0521821509.

Mohler, R.R. (1973), Bilinear Control Processes with Applications to Engineering, Ecology and Medicine. Math. Sci. Engrg. 106.

Pemberton, J. (1993), On Tests for Non-Linearity in Time Series: Paper Presented at the International Conference on Applications of Time Series Analysis in Astronomy and Meteology, University of Padua, Italy, September: 6-10.

Pemberton, J. (1994), A Test for Martingale Difference Errors in Time Series based on the Petruccelli-Davies Procedure. Working Paper EBMS.

Petruccelli, J.D. and Davies, N. (1986), A Portmanteau Test for Self-Exciting Threshold Autoregressive Type Non-Linearity in Time Series. Bimetrika., 73:687-94.

Priestly, M.B. (1980), State Dependent Models: A General Approach to Non-Linear Time Series. J. Time Series Anal.1: 47-71.

Priestly, M.B. (1981), Spectral Analysis and Time Series. Univariate Time Series. Academic Press London,1.

Rothman, P. (1999), Non Linear Time Series Analysis of Economic and Financial Data. Kluwer Academic Publishers., ISBN 0792383796.

Ruberti, A., Isidori, A. and D'Alessandro, P. (1972), Theory of Bilinear Dynamical Systems, Springer, Berlino.

Tong, H. (1978), On a Threshold Model. In Pattern Recognition and Signal Processing, (ed. C.H.Chen). Sijthoff and Noordhoff, Amsterdam.

Tong, H. (1983), Threshold Models in Non-Linear Time Series Analysis, Lecture Notes in Statistics, 21, (ed.D. Brillinger, S. Fienberg, J. Gani, J. Hartigan and K.Krickeberry). Springer-Verlag, New York.

Tong, H. and Lim, K.S. (1980), Threshold Autoregression, Limit Cycles and Cyclical Data (with discussion). J.R. Statist. Soc. B42, 245-92.

Tong. H. (1990), Non-Linear Time Series: A Dynamical System Approach. Oxford Science Publications.

Yule, G.U. (1927), On a Method of Investigating Periodicities in Disturbed Series with Spectral Reference to Wolfer's Sunspot Numbers. Philos Trans. Roy. Soc., London, Series A., 226: 367-98.

APPENDIX A

Conditional Distribution in the Laguerre Mixture Model

The following derivation of the t-step conditional distribution in the Laguerre model was kindly provided in hand written notes by Jalali (1993).

Considering the Laguerre model with weights, $\alpha_i(z)=(\theta z)^i \exp(-\theta z)/i!$ and densities, $f_i(z)=\lambda(\lambda z)^i \exp(-\lambda z)/i!$ for i=0,1,2,...; $z\geq 0$, in Proposition (4.3.2) we showed the transition matrix $\boldsymbol{P}$ has the following entries

$$p_{ij}=\int_{-\infty}^{\infty}\alpha_j(z)f_i(z)dz$$

$$=\int_{-\infty}^{\infty}\frac{(\theta z)^j \exp(-\theta z)}{j!}\frac{\lambda(\lambda z)^i \exp(-\lambda z)}{i!}dz$$

$$=\frac{\lambda^{i+1}\theta^j}{i!j!}\left[\int_{-\infty}^{\infty}\frac{(\lambda+\theta)^{i+j+1}z^{i+j}\exp\left(-(\lambda+\theta)z\right)}{(i+j)!}dz\right]\frac{(i+j)!}{(\lambda+\theta)^{i+j+1}}$$

$$=\frac{\lambda}{\lambda+\theta}\frac{(i+j)!}{i!j!}\left(\frac{\lambda}{\lambda+\theta}\right)^i\left(\frac{\theta}{\lambda+\theta}\right)^j.$$

Setting, $q=\frac{\lambda}{\lambda+\theta}$ and $s=\frac{\theta}{\lambda+\theta}$ i.e. $r=\frac{s}{q}$ we have

$$p_{ij}=q\binom{i+j}{i}q^i s^j. \qquad \text{(A.1)}$$

Let us have a vector of weights on non-negative integers. Clearly this vector has a generating function, say, $g(n)$ defined on, and of the boundary of, the unit disc where $g(n)=\sum_{i=0}^{\infty}q_i n^i$. Now, let $\boldsymbol{P}$ act on this vector of weights and we want to see what is the generating function, say, $\boldsymbol{P}(g)$ of this new vector of weights.

$$P(g(n)) = \sum_{j=0}^{\infty} \left(\sum_{i=0}^{\infty} q_i p_{ij} \right) n^j$$

$$= \sum_{i=0}^{\infty} q_i \sum_{j=0}^{\infty} q \binom{i+j}{j} q^i s^j n^j$$

$$= q \sum_{i=0}^{\infty} q_i q^i \sum_{j=0}^{\infty} \binom{-i-1}{j} (-1)^j s^j n^j$$

$$= q \sum_{i=0}^{\infty} q_i q^i \frac{1}{(1-sn)^{i+1}}$$

$$= \frac{q}{1-sn} \sum_{i=0}^{\infty} q_i \left(\frac{q}{1-sn} \right)^i$$

$$= \frac{q}{1-sn} g\left(\frac{q}{1-sn} \right) \qquad \text{(A.2)}$$

Since our initial weights are $(\theta z)^i \exp(-\theta z) / i!$, the corresponding generating function is

$$g(n) = \sum_{i=0}^{\infty} \frac{(\theta z)^i}{i!} \exp(-\theta z) n^i$$

$$= \exp(-\theta z) \sum_{i=0}^{\infty} \frac{(\theta z n)^i}{i!}$$

$$= \exp(\theta z (n-1)) \qquad \text{(A.3)}$$

which is defined over all the complex plane.

Setting $g_0=g$, $g_t=P^t(g)$ and $h_t=\ln g_t$ we have

$$h_0(n)=\ln g_0(n)=\ln g(n)=\ln\left(\exp(\theta z(n-1))\right)=\theta z(n-1)$$

$$h_{t+1}(n)=\ln g_{t+1}(n)=\ln\left[P\left(P^t(g(n))\right)\right]=\ln\left[\left(\frac{q}{1-sn}\right)g_t\left(\frac{q}{1-sn}\right)\right]=\ln\left(\frac{q}{1-sn}\right)+h_t\left(\frac{q}{1-sn}\right).$$

Now we let $w(n)=\frac{q}{1-sn}$, $\alpha=\theta z$ and so

$$h_0(n)=\alpha(n-1)$$

$$h_{t+1}(n)=\ln w(n)+h_t(w(n)).$$

Therefore $h_1(n)=\ln w(n)+h_0(w(n))=\ln w(n)+\alpha(w(n)-1)$

$$h_2(n)=\ln w(n)+h_1(w(n))=\ln w(n)+\ln w^2(n)+\alpha\left(w^2(n)-1\right)$$

$$\vdots$$

$$h_t(n)=\ln w(n)+\ln w^2(n)+\ldots+\ln w^t(n)+\alpha\left(w^t(n)-1\right).$$

Let us now denote the function $w(n)=\frac{q}{1-sn}$ by the matrix $\begin{pmatrix}0 & q\\ -s & 1\end{pmatrix}$ and so

w^t corresponds to $\begin{pmatrix}0 & q\\ -s & 1\end{pmatrix}^t$.

Since we defined in (5.2) that r<1, it follows $q\neq s$. Therefore, $\begin{pmatrix}0 & q\\ -s & 1\end{pmatrix}^t=q^tA+s^tB$. To find A and B we note that

$$\begin{pmatrix}0 & q\\ -s & 1\end{pmatrix}^0=q^0A+s^0B\Rightarrow A+B=I \text{ and}$$

$$\begin{pmatrix}0 & q\\ -s & 1\end{pmatrix}^1=qA+sB\Rightarrow\begin{pmatrix}0 & q\\ -s & 1\end{pmatrix}=qA+s(I-A).$$

Hence $A=\begin{pmatrix}-s & q\\ -s & q\end{pmatrix}\Big/(q-s)$ and $B=\begin{pmatrix}q & -q\\ s & -s\end{pmatrix}\Big/(q-s)$.

Therefore, $\begin{pmatrix}0 & q\\ -s & 1\end{pmatrix}^t=\frac{1}{q-s}\left(q^t\begin{pmatrix}-s & q\\ -s & q\end{pmatrix}+s^t\begin{pmatrix}q & -q\\ s & -s\end{pmatrix}\right)$

$$=\frac{1}{q-s}\begin{pmatrix}-qs\left(q^{t-1}-s^{t-1}\right) & q\left(q^t-s^t\right)\\ -s\left(q^t-s^t\right) & q^{t+1}-s^{t+1}\end{pmatrix}$$

and $w^t(n)=\dfrac{-qs\left(q^{t-1}-s^{t-1}\right)n+q\left(q^t-s^t\right)}{-s\left(q^t-s^t\right)n+q^{t+1}-s^{t+1}}$

$$=\frac{-\left(r-r^t\right)n+\left(1-r^t\right)}{-\left(r-r^{t+1}\right)n+\left(1-r^{t+1}\right)}. \qquad \text{(A.4)}$$

Now for an arbitrary t we have

$$g_t(n)=w(n)w^2(n)\ldots w^t(n)\exp\left(\alpha\left(w^t(n)-1\right)\right)$$

$$=\frac{1-r}{-\left(r-r^2\right)n+\left(1-r^2\right)}\frac{-\left(r-r^2\right)n+\left(1-r^2\right)}{-\left(r-r^3\right)n+\left(1-r^3\right)}\cdots\frac{-\left(r-r^t\right)n+\left(1-r^t\right)}{-\left(r-r^{t+1}\right)n+\left(1-r^{t+1}\right)}$$

$$\exp\left(\alpha\left(\frac{-\left(r-r^t\right)n+\left(1-r^t\right)}{-\left(r-r^{t+1}\right)n+\left(1-r^{t+1}\right)}-1\right)\right)$$

$$=\frac{1-r}{\left(1-r^{t+1}\right)-r\left(1-r^t\right)n}\exp\left(-\alpha r^t(1-r)\frac{1-n}{\left(1-r^{t+1}\right)-r\left(1-r^t\right)n}\right)$$

$$=\frac{1-r}{1-r^{t+1}}\frac{1}{1-\dfrac{r\left(1-r^t\right)n}{1-r^{t+1}}}\exp\left(\frac{-\alpha r^t(1-r)}{1-r^{t+1}}\frac{1-n}{1-\dfrac{r\left(1-r^t\right)n}{1-r^{t+1}}}\right). \qquad \text{(A.5)}$$

Setting $u_{t+1}=\frac{1-r}{1-r^{t+1}}$, $v_{t+1}=\frac{r\left(1-r^{t}\right)}{1-r^{t+1}}$ we have

$$g_t(n)=\frac{u_{t+1}}{1-v_{t+1}n}\exp\left(-\alpha r^t u_{t+1}\frac{1-n}{1-v_{t+1}n}\right)$$

$$=\frac{u_{t+1}}{1-v_{t+1}n}\exp\left(-\alpha r^t u_{t+1}\right)\exp\left(\alpha r^t\frac{u_{t+1}^2 n}{1-v_{t+1}n}\right). \qquad \text{(A.6)}$$

The generating function of the Laguerre model is $(1-w)^{-1}\exp\left(\frac{-w\bar{z}}{1-w}\right)=\sum_{m=0}^{\infty}L_m(\bar{z})w^m$,

where $L_m(\bar{z})$ is the m^{th} order Laguerre polynomial.

Let $w=v_{t+1}n$, $w\bar{z}=-\alpha r^t u_{t+1}^2 n$, hence

$$g_t(n)=\frac{u_{t+1}}{1-w}\exp\left(-\alpha r^t u_{t+1}\right)\exp\left(\frac{-w\bar{z}}{1-w}\right)$$

$$=u_{t+1}\exp\left(-\alpha r^t u_{t+1}\right)\sum_{m=0}^{\infty}L_m\left(-\alpha r^t\frac{u_{t+1}^2}{v_{t+1}}\right)v_{t+1}^m n^m. \qquad \text{(A.7)}$$

If we start with the vector of weights $\alpha_m^{(0)}(z)=(\theta z)^m\exp(-\theta z)/m!$ then after t-times transformation by $\boldsymbol{P}$ we obtain

$\alpha_m^{(t)}(z)$ =coefficient of n^m in $g_t(n)$

$$=u_{t+1}\exp\left(-\alpha r^t u_{t+1}\right)L_m\left(-\alpha r^t\frac{u_{t+1}^2}{v_{t+1}}\right)v_{t+1}^m$$

$$=u_{t+1}v_{t+1}^m\exp\left(-\theta r^t u_{t+1}z\right)L_m\left(-\theta r^t\frac{u_{t+1}^2}{v_{t+1}}z\right). \qquad \text{(A.8)}$$

When $t\to\infty$, $u_{t+1}\to(1-r)$, $v_{t+1}\to r$ and thus $\alpha_m^{(\infty)}(z)=(1-r)r^m L_m(0)$.

Note that $L_m(0)=\sum_{k=0}^{m}\binom{m}{m-k}\frac{(-0)^k}{k!}=1$ and so $\alpha_m^{(\infty)}(z)=(1-r)r^m$ for m=0,1,2,... which is the p.d.f. of a geometric distribution.

To find the $\alpha_m^{(t)}$ mixture of our f_m's we need to find the sum of the following infinite series

$$\sum_{m=0}^{\infty} L_m(x)\frac{w^m}{m!}.$$

Therefore, once more we invoke the formula $L_m(x)=\sum_{k=0}^{m}\binom{m}{m-k}\frac{(-x)^k}{k!}$.

$$\begin{aligned}\sum_{m=0}^{\infty} L_m(x)\frac{w^m}{m!} &= \sum_{m=0}^{\infty}\sum_{k=0}^{m}\binom{m}{m-k}\frac{(-x)^k}{k!}\frac{w^m}{m!}\\ &= \sum_{k=0}^{\infty}\sum_{m=k}^{\infty}\frac{w^m}{(m-k)!(k!)^2}(-x)^k\\ &= \sum_{k=0}^{\infty}\frac{(-x)^k w^k}{(k!)^2}\sum_{m-k=0}^{\infty}\frac{w^{m-k}}{(m-k)!}\\ &= \exp(w)\sum_{k=0}^{\infty}\frac{(-xw)^k}{(k!)^2} \qquad \text{(A.9)}\end{aligned}$$

It is well known that the modified bessel function of order zero $I_0(x)$, has the following form

$$I_0(x)=\sum_{m=0}^{\infty}\frac{\left(\left(\frac{x}{2}\right)^2\right)^m}{(m!)^2}.$$

Hence if xw is negative then $\sum_{k=0}^{\infty}\frac{(-xw)^k}{(k!)^2}=\sum_{k=0}^{\infty}\frac{\left(\frac{2(-xw)^{1/2}}{2}\right)^{2k}}{(k!)^2}=I_0\left(2\sqrt{(-xw)}\right)$

and so $\sum_{m=0}^{\infty} L_m(x)\frac{w^m}{m!}=\exp(w)I_0\left(2\sqrt{(-xw)}\right)$.

The conditional distribution of Z_{t+1} given Z_0 has the p.d.f. as follows;

$$f(z_{t+1}|z_0) = \sum_{m=0}^{\infty} \alpha_m^{(t)}(z_0) f_m(z_{t+1})$$

$$= \sum_{m=0}^{\infty} u_{t+1} v_{t+1}^m \exp\left(-\theta r^t u_{t+1} z_0\right) L_m\left(-\theta r^t \frac{u_{t+1}}{v_{t+1}} z_0\right)$$

$$\lambda(\lambda z_{t+1})^m \exp(-\lambda z_{t+1}) / m!$$

$$= \lambda u_{t+1} \exp\left(-\theta r^t u_{t+1} z_0\right) \exp(-\lambda z_{t+1}) \sum_{m=0}^{\infty} \frac{\lambda^m v_{t+1}^m z_{t+1}^m}{m!} L_m\left(-\theta r^t \frac{u_{t+1}}{v_{t+1}} z_0\right)$$

$$= \lambda u_{t+1} \exp\left(-\theta r^t u_{t+1} z_0\right) \exp(-\lambda u_{t+1} z_{t+1}) I_0\left(2\sqrt{\left(\theta r^t \lambda u_{t+1}^2 z_0 z_{t+1}\right)}\right).$$

Replacing θ by $r\lambda$ gives us

$$f(z_{t+1}|z_0) = \lambda u_{t+1} \exp\left(-\lambda u_{t+1}\left(z_{t+1} + z_0 r^{t+1}\right)\right) I_0\left(2\lambda u_{t+1}\sqrt{\left(z_0 z_{t+1} r^{t+1}\right)}\right).$$

Hence,

$$f(z_t|z_0) = \lambda u_t \exp\left(-\lambda u_t\left(z_t + z_0 r^t\right)\right) I_0\left(2\lambda u_t \sqrt{\left(z_0 z_t r^t\right)}\right). \qquad \text{(A.10)}$$

APPENDIX B

Log-likelihood of Laguerre Mixture Model

This program simulates a time series of length t+1=500 data observations from the Laguerre model with true values r=0.5 and λ=10.0. To create a series with missing values, a certain proportion of the data observations $z_1, z_2, \ldots, z_{t-1}$ are replaced at random for zero's. Here, a value of δ=0.1 is used and hence, 10% of the observations are replaced with zero. The program computes the log-likelihood function over the range $r \in [0.1, 0.9]$ and $\lambda \in [5.0, 15.0]$.

```
      PROGRAM LOGLIKELIHOOD
C     Generates a time series from the Laguerre mixture model and calculates
C     the log-likelihood function over the range of r and lambda.
      REAL*8 rr,YY,LL
      INTEGER IFAIL
C
      REAL*8 RAIN(1000)
      REAL*8 trueY,truer,delta
      INTEGER K,NZERO
C
      COMMON/BLOCK1/RAIN,K
      OPEN(5,FILE='LOGLIKELIHOOD.DAT',STATUS='NEW')
      CALL G05CBF(0)
C
      K=500
      trueY=10.0D0
      truer=0.5D0
      delta=0.1D0
      NZERO=delta*k
      CALL DATA(trueY,truer,NZERO)
      DO 50 YY=5.0D0,15.0D0,0.5D0
        DO 60 rr=0.1D0,0.9D0,0.1D0
          CALL FUNCT(rr,YY,LL)
          WRITE(5,100) rr,YY,LL
60      CONTINUE
50    CONTINUE
100   FORMAT('[',F20.10,',',F20.10,',',F20.10,']',',')
      STOP
      END
C-----------------------------------------------------------------------
      SUBROUTINE DATA(trueY,truer,NZERO)
C     Simulates data of length K from Laguerre(r,Y).
      REAL*8 RAIN(1000)
      REAL*8 trueY,truer,C,D,E,XD,Z
      REAL*8 G05CAF,G05DBF,G05DGF,R(1000)
      INTEGER MISS(1000)
      INTEGER I,J,K,NZERO,IY,G05EYF
C
C     G05CAF returns a pseudo-random real number taken from a
C     uniform distribution between 0 and 1.
C     G05DBF returns a psuedo-random real number taken from an
C     exponential distribution with mean (1/(Y*(1-r))).
```

```
C       G05ECF sets up the reference vector R for a poisson distribution
C       with mean D.
C       G05DGF returns a pseudo-random real number taken from a gamma
C       distribution with parameters E and (1/Y).
C       G05CBF sets the generator routine to a repeatable initial state.
C
        COMMON/BLOCK1/RAIN,K
C
        DO 10 I=1,K
          MISS(I)=1
10      CONTINUE
        DO 20 I=1,NZERO
98        Z=G05CAF(Z)
          IF(Z .LT. 1.0D0/DFLOAT(K) .OR.
     *    Z .GE. (DFLOAT(K)-1.0D0)/DFLOAT(K))THEN
            GOTO 98
          ENDIF
          DO 30 J=2,K-1
            IF(Z .GE. DFLOAT(J-1)/DFLOAT(K)
     *      .AND. Z .LT. DFLOAT(J)/DFLOAT(K))THEN
              IF(MISS(J) .EQ. 1)THEN
                MISS(J)=0
              ELSE
                GOTO 98
              ENDIF
            ENDIF
30        CONTINUE
20      CONTINUE
        XD=G05DBF(1.0D0/(trueY*(1.0D0-truer)))
        RAIN(1)=XD
        IFAIL=0
        C=trueY*truer
        D=C*RAIN(1)
        DO 40 I=2,K
          CALL G05ECF(D,R,1000,IFAIL)
          IY=G05EYF(R,1000)
          E=dble(IY)+1.0D0
          RAIN(I)=G05DGF(E,(1.0D0/trueY),IFAIL)
          D=C*RAIN(I)
            IF(MISS(I) .EQ. 0)THEN
              RAIN(I)=0.0D0
            ENDIF
40      CONTINUE
        RETURN
        END
C-----------------------------------------------------------------------
        SUBROUTINE FUNCT(rr,YY,LL)
C       Function evaluation routine for calculation of maximum likelihood.
        REAL*8 U(1000),RAIN(1000)
        REAL*8 XX,ZZ,rr,YY,LL,S18AEF
        INTEGER S(0:1000),T(1000)
        INTEGER A,I,J,K,L,M,Num,IFAIL
C
C       S18AEF returns the value of the modified bessel function I.(X),via
C       the routine name.
C
        COMMON/BLOCK1/RAIN,K
C
        LL=0.0D0
        S(0)=1
        L=1
```

```
      A=0
      DO 10 I=2,K
        IF(RAIN(I) .GT. 0.0D0)THEN
          S(L)=I
          T(L)=S(L)-S(L-1)
          IF(A .LT. T(L))THEN
            A=T(L)
          ENDIF
          L=L+1
        ENDIF
10    CONTINUE
      DO 20 M=1,A
        U(M)=(1.0D0-rr)/(1.0D0-rr**M)
20    CONTINUE
      I=1
      Num=1
      DO 30 J=2,K
        IF(RAIN(J) .GT. 0.0D0)THEN
          IFAIL=-1
          XX=2.0D0*YY*U(T(Num))*SQRT(RAIN(I)*RAIN(J)*(rr**T(Num)))
          IF(XX .GT. 87.9D0)THEN
            CALL POLY(XX,ZZ)
          ELSE
            ZZ=S18AEF(XX,IFAIL)
            ZZ=DLOG(ZZ)
            IF(IFAIL .NE. 0)THEN
              PRINT*,'ERROR IN NAG:IFAIL=',IFAIL
              STOP
            ENDIF
          ENDIF
          LL=LL+DLOG(YY)+DLOG(U(T(Num)))-YY*U(T(Num))
     *    *(RAIN(J)+RAIN(I)*(rr**T(Num)))+(ZZ)
          Num=Num+1
          I=J
        ENDIF
30    CONTINUE
      LL=LL+DLOG(YY)+DLOG(1.0D0-rr)-(YY*(1.0D0-rr)*RAIN(1))
      RETURN
C     End of function evaluation routine.
      END
C-----------------------------------------------------------------------
      SUBROUTINE POLY(XX,ZZ)
C     Polynomial approximation of I.(X) when the value of X
C     exceeds 87.9 where the nag routine can not compute.
      REAL*8 XX,ZZ,t
C
      t=3.75D0/XX
      ZZ=0.39894228D0+(0.01328592D0*t)+(0.00225319D0*(t**2))
     *  -(0.00157565D0*(t**3))+(0.00916281D0*(t**4))
     *  -(0.02057706D0*(t**5))+(0.02635537D0*(t**6))
     *  -(0.01647633D0*(t**7))+(0.00392377D0*(t**8))
      ZZ=DLOG(ZZ)-(0.5D0*DLOG(XX))+XX
      RETURN
      END
```

APPENDIX C

Hessian Matrix of Laguerre Mixture Model

This program simulates a time series of length t+1=500 data observations from the Laguerre model with true values r=0.5 and λ=10.0. A value of δ=0.1 is used again and hence, 10% of the data observations $z_1, z_2, \ldots, z_{t-1}$ are replaced at random for zero's. The program computes the determinant of the Hessian matrix over the range $r \in [0.1, 0.9]$ and $\lambda \in [5.0, 15.0]$.

```
      PROGRAM DETERMINANT
C     Generates a time series from the Laguerre mixture model and
C     evaluates the determinant of the Hessian matrix over the
C     range of r and lambda.
      REAL*8 rr,YY,DET
      INTEGER IFAIL
C
      REAL*8 RAIN(1000)
      REAL*8 trueY,truer,delta
      INTEGER K,NZERO
C
      COMMON/BLOCK1/RAIN,K
      OPEN(5,FILE='DETERMINANT.DAT',STATUS='NEW')
      CALL G05CBF(0)
C
      K=500
      trueY=10.0D0
      truer=0.5D0
      delta=0.1D0
      NZERO=delta*k
      CALL DATA(trueY,truer,NZERO)
      DO 50 YY=5.0D0,15.0D0,0.1D0
        DO 60 rr=0.1D0,0.9D0,0.1D0
          CALL FUNCT(rr,YY,DET)
          WRITE(5,100) rr,YY,DET
60      CONTINUE
50    CONTINUE
100   FORMAT('[',F20.10,',',F20.10,',',F20.10,']',',')
      STOP
      END
C-----------------------------------------------------------------
      SUBROUTINE DATA(trueY,truer,NZERO)
C     Simulates data of length K from Laguerre(r,Y).
      REAL*8 RAIN(1000)
      REAL*8 trueY,truer,C,D,E,XD,Z
      REAL*8 G05CAF,G05DBF,G05DGF,R(1000)
      INTEGER MISS(1000)
      INTEGER I,J,K,NZERO,IY,G05EYF
C
C     G05CAF returns a pseudo-random real number taken from a
C     uniform distribution between 0 and 1.
C     G05DBF returns a psuedo-random real number taken from an
C     exponential distribution with mean (1/(Y*(1-r))).
C     G05ECF sets up the reference vector R for a poisson distribution
```

```
C       with mean D.
C       G05DGF returns a pseudo-random real number taken from a gamma
C       distribution with parameters E and (1/Y).
C       G05CBF sets the generator routine to a repeatable initial state.
C
        COMMON/BLOCK1/RAIN,K
C
        DO 10 I=1,K
          MISS(I)=1
10      CONTINUE
        DO 20 I=1,NZERO
98        Z=G05CAF(Z)
          IF(Z .LT. 1.0D0/DFLOAT(K) .OR.
     *    Z .GE. (DFLOAT(K)-1.0D0)/DFLOAT(K))THEN
            GOTO 98
          ENDIF
          DO 30 J=2,K-1
            IF(Z .GE. DFLOAT(J-1)/DFLOAT(K)
     *      .AND. Z .LT. DFLOAT(J)/DFLOAT(K))THEN
              IF(MISS(J) .EQ. 1)THEN
                MISS(J)=0
              ELSE
                GOTO 98
              ENDIF
            ENDIF
30        CONTINUE
20      CONTINUE
        XD=G05DBF(1.0D0/(trueY*(1.0D0-truer)))
        RAIN(1)=XD
        IFAIL=0
        C=trueY*truer
        D=C*RAIN(1)
        DO 40 I=2,K
          CALL G05ECF(D,R,1000,IFAIL)
          IY=G05EYF(R,1000)
          E=dble(IY)+1.0D0
          RAIN(I)=G05DGF(E,(1.0D0/trueY),IFAIL)
          D=C*RAIN(I)
          IF(MISS(I) .EQ. 0)THEN
            RAIN(I)=0.0D0
          ENDIF
40      CONTINUE
        RETURN
        END
C-----------------------------------------------------------------------
        SUBROUTINE FUNCT(rr,YY,DET)
C       Function evaluation routine for calculation of determinant.
        REAL*8 U(1000),UD(1000),UDD(1000),RAIN(1000)
        REAL*8 XX,ZZ,WW,rr,YY,DET,S18AEF,S18AFF
        REAL*8 B,Br,BY,dr2,dY2,drY
        INTEGER S(0:1000),T(1000)
        INTEGER A,I,J,K,L,M,Num,IFAIL
C
C       S18AEF returns the value of the modified bessel function I.(X),via
C       the routine name.
C       S18AFF returns the value of the modified bessel function I,(X),via
C       the routine name.
C
        COMMON/BLOCK1/RAIN,K
C
        S(0)=1
```

```
      L=1
      A=0
      DO 10 I=2,K
        IF(RAIN(I) .GT. 0.0D0)THEN
          S(L)=I
          T(L)=S(L)-S(L-1)
          IF(A .LT. T(L))THEN
            A=T(L)
          ENDIF
          L=L+1
        ENDIF
10    CONTINUE
      DO 20 M=1,A
        U(M)=(1.0D0-rr)/(1.0D0-rr**M)
        UD(M)=-U(M)*(1.0D0-U(M)*M*(rr**(M-1)))/(1.0D0-rr)
        UDD(M)=M*(rr**(M-2))*((M-1)*U(M)+2.0D0*rr*UD(M))
     *  /(1.0D0-(rr**M))
20    CONTINUE
      I=1
      Num=1
      Dr2=0.0D0
      DY2=0.0D0
      DrY=0.0D0
      DET=0.0D0
      DO 30 J=2,K
        IF(RAIN(J) .GT. 0.0D0)THEN
          IFAIL=-1
          XX=2.0D0*YY*U(T(Num))*SQRT(RAIN(I)*RAIN(J)*(rr**T(Num)))
          IF(XX .GT. 87.9D0)THEN
            CALL POLY1(XX,ZZ)
            CALL POLY2(XX,WW)
          ELSE
            ZZ=S18AEF(XX,IFAIL)
            ZZ=DLOG(ZZ)
            WW=S18AFF(XX,IFAIL)
            WW=DLOG(WW)
            IF(IFAIL .NE. 0)THEN
              PRINT*,'ERROR IN NAG:IFAIL=',IFAIL
              STOP
            ENDIF
          ENDIF
          B=XX*DEXP(WW)/DEXP(ZZ)
          Br=((T(NUM)/(2.0D0*rr))+(UD(T(NUM))/U(T(NUM))))
     *    *(B+(XX**2)*((DEXP(ZZ)**2)-(DEXP(ZZ)*DEXP(WW))/XX
     *    -(DEXP(WW)**2))/(DEXP(ZZ)**2))
          Y=(1.0D0/YY)*(B+(XX**2)*((DEXP(ZZ)**2)-(DEXP(ZZ)
     *    *DEXP(WW))/XX-(DEXP(WW)**2))/(DEXP(ZZ)**2))
          dr2=dr2-(YY*U(T(NUM))*RAIN(I)*T(NUM)*(T(NUM)-1)
     *    *(rr**(T(NUM)-2)))
     *    -(2.0D0*YY*RAIN(I)*T(NUM)*(rr**(T(NUM)-1))+UD(T(NUM))/
     *    (U(T(NUM))**2))*UD(T(NUM))
     *    -(YY*(RAIN(J)+RAIN(I)*(rr**T(NUM)))-(1.0D0/U(T(NUM))))
     *    *UDD(T(NUM))
     *    -(T(NUM)/(2.0D0*(rr**2))+(UD(T(NUM))**2)/(U(T(NUM))**2)
     *    -(UDD(T(NUM))/U(T(NUM))))*B
     *    +(T(NUM)/(2.0D0*rr)+(UD(T(NUM))/U(T(NUM))))*Br
          dY2=dY2-1.0D0/(YY**2)-B/(YY**2)+BY/YY
          drY=drY-U(T(NUM))*RAIN(I)*T(NUM)*(rr**(T(NUM)-1))
     *    -(RAIN(J)+RAIN(I)*(rr**T(NUM)))*UD(T(NUM))+Br/YY
          Num=Num+1
          I=J
```

```
          ENDIF
30        CONTINUE
          dr2=dr2-(1.0D0/((1.0D0-rr)**2))
          dY2=dY2-(1.0D0/(YY**2))
          drY=drY+RAIN(1)
          DET=dr2*dY2-(drY**2)
          RETURN
C         End of function evaluation routine.
          END
C-------------------------------------------------------------------------
          SUBROUTINE POLY1(XX,ZZ)
C         Polynomial approximation of I.(X) when the value of X
C         exceeds 87.9 where the nag routine can not compute.
          REAL*8 XX,ZZ,t
C
          t=3.75D0/XX
          ZZ=0.39894228D0+(0.01328592D0*t)+(0.00225319D0*(t**2))
     *    -(0.00157565D0*(t**3))+(0.00916281D0*(t**4))
     *    -(0.02057706D0*(t**5))+(0.02635537D0*(t**6))
     *    -(0.01647633D0*(t**7))+(0.00392377D0*(t**8))
          ZZ=DLOG(ZZ)-(0.5D0*DLOG(XX))+XX
          RETURN
          END
C-------------------------------------------------------------------------
          SUBROUTINE POLY2(XX,WW)
C         Polynomial approximation of I,(X) when the value of X
C         exceeds 87.9 where the nag routine can not compute.
          REAL*8 XX,WW,t
C
          t=3.75D0/XX
          WW=0.39894228D0-(0.03988024D0*t)-(0.00362018D0*(t**2))
     *    +(0.00163801D0*(t**3))-(0.01031555D0*(t**4))
     *    +(0.02282967D0*(t**5))-(0.02895312D0*(t**6))
     *    +(0.01787654D0*(t**7))-(0.00420059D0*(t**8))
          WW=DLOG(WW)-(0.5D0*DLOG(XX))+XX
          RETURN
          END
```

APPENDIX D

Maximum Likelihood Estimation of Laguerre Mixture Model

This program simulates 100 time series of lengths 50, 250, 500 and 1000 data observations from the Laguerre model using a range of values for true r and λ. A value of δ=0.1 is used again and hence, 10% of the data observations $z_1, z_2, \ldots, z_{t-1}$ are replaced at random for zero's. The maximisation routine E04JAF is then applied to the log-likelihood function using this data and maximum likelihood estimates of r, λ are given.

```
      PROGRAM MAXLIKELIHOOD
C     Generates a time series from the Laguerre mixture model and
C     calculates maximum likelihood estimates of r and lambda.
      REAL*8 F
      REAL*8 BL(2),BU(2),W(25),X(2)
      INTEGER IBOUND,IFAIL,Q,LIW,LW,B
      INTEGER IW(4)
C
      REAL*8 RAIN(1000)
      REAL*8 trueY,truer,delta
      INTEGER K,N,NZERO,KOUNT
C
      COMMON/BLOCK1/RAIN,K
      OPEN(5,FILE='MAXLIKELIHOOD.DAT',STATUS='NEW')
      CALL G05CBF(0)
C
      PRINT*,'Input - no. of iterations,size of data,true Y,true r,delta'
      READ*,N,K,trueY,truer,delta
      NZERO=delta*K
      KOUNT=0
97    IF(KOUNT .LT. N)THEN
        CALL DATA(trueY,truer,NZERO)
C
        B=2
        X(1)=truer
        X(2)=trueY
        IBOUND=0
        BL(1)=0.00001D0
        BU(1)=0.99999D0
        BL(2)=0.1D0
        BU(2)=100.0D0
        LIW=4
        LW=25
        IFAIL=-1
        CALL E04JAF(B,IBOUND,BL,BU,X,F,IW,LIW,W,LW,IFAIL)

C       E04JAF is a quasi-newton algorithm for finding the minimum of a
C       function F(X1,X2),subject to fixed upper and lower bounds on the
C       independant variables X1,X2, using function values only.
C       Since IFAIL was set to -1 before entering E04JAF,it is
C       essential to test whether IFAIL is NON-ZERO on EXIT.

        IF(IFAIL .NE. 0)THEN
          GOTO 97
```

```
          ENDIF
          IF(IFAIL .EQ. 1)GOTO 99
          PRINT100,F
          PRINT200,(X(Q),Q=1,B)
          WRITE(5,300)(X(Q),Q=1,B)
          KOUNT=KOUNT+1
          GOTO 97
        ENDIF
99      STOP
100     FORMAT(4H LL=,F30.10)
200     FORMAT(4H r =,F30.10,5H   Y=,F30.10)
300     FORMAT(2F30.10)
        END
C-------------------------------------------------------------------
        SUBROUTINE DATA(trueY,truer,NZERO)
C       Simulates data of length K from Laguerre(r,Y).
        REAL*8 RAIN(1000)
        REAL*8 trueY,truer,C,D,E,XD,Z
        REAL*8 G05CAF,G05DBF,G05DGF,R(1000)
        INTEGER MISS(1000)
        INTEGER I,J,K,NZERO,IY,G05EYF
C
C       G05CAF returns a pseudo-random real number taken from a
C       uniform distribution between 0 and 1.
C       G05DBF returns a pseudo-random real number taken from an
C       exponential distribution with mean (1/(Y*(1-r))).
C       G05ECF sets up the reference vector R for a poisson distribution
C       with mean D.
C       G05DGF returns a pseudo-random real number taken from a gamma
C       distribution with parameters E and (1/Y).
C       G05CBF sets the generator routine to a repeatable initial state.
C
        COMMON/BLOCK1/RAIN,K
C
        DO 10 I=1,K
          MISS(I)=1
10      CONTINUE
        DO 20 I=1,NZERO
98        Z=G05CAF(Z)
          IF(Z .LT. 1.0D0/DFLOAT(K) .OR.
     *    Z .GE. (DFLOAT(K)-1.0D0)/DFLOAT(K))THEN
            GOTO 98
          ENDIF
          DO 30 J=2,K-1
            IF(Z .GE. DFLOAT(J-1)/DFLOAT(K)
     *       .AND. Z .LT. DFLOAT(J)/DFLOAT(K))THEN
              IF(MISS(J) .EQ. 1)THEN
                MISS(J)=0
              ELSE
                GOTO 98
              ENDIF
            ENDIF
30        CONTINUE
20      CONTINUE
        XD=G05DBF(1.0D0/(trueY*(1.0D0-truer)))
        RAIN(1)=XD
        IFAIL=0
        C=trueY*truer
        D=C*RAIN(1)
        DO 40 I=2,K
          CALL G05ECF(D,R,1000,IFAIL)
```

```
          IY=G05EYF(R,1000)
          E=dble(IY)+1.0D0
          RAIN(I)=G05DGF(E,(1.0D0/trueY),IFAIL)
          D=C*RAIN(I)
          IF(MISS(I) .EQ. 0)THEN
            RAIN(I)=0.0D0
          ENDIF
40        CONTINUE
          RETURN
          END
C-----------------------------------------------------------------------
          SUBROUTINE FUNCT1(B,XC,FC)
C         Function evaluation routine for E04JAF.
          REAL*8 FC
          INTEGER B
          REAL*8 XC(B)
C
          REAL*8 U(1000),RAIN(1000)
          REAL*8 XX,ZZ,r,Y,S18AEF
          INTEGER S(0:1000),T(1000)
          INTEGER A,I,J,K,L,M,Num,IFAIL
C
C         S18AEF returns the value of the modified bessel function I.(X),via
C         the routine name.
C
          COMMON/BLOCK1/RAIN,K
C
          r=XC(1)
          Y=XC(2)
          FC=0.0D0
          S(0)=1
          L=1
          A=0
          DO 10 I=2,K
            IF(RAIN(I) .GT. 0.0D0)THEN
              S(L)=I
              T(L)=S(L)-S(L-1)
              IF(A .LT. T(L))THEN
                A=T(L)
              ENDIF
              L=L+1
            ENDIF
10        CONTINUE
          DO 20 M=1,A
            U(M)=(1.0D0-r)/(1.0D0-r**M)
20        CONTINUE
          I=1
          Num=1
          DO 30 J=2,K
            IF(RAIN(J) .GT. 0.0D0)THEN
              IFAIL=-1
              XX=2.0D0*Y*U(T(Num))*SQRT(RAIN(I)*RAIN(J)*(r**T(Num)))
              IF(XX .GT. 87.9D0)THEN
                CALL POLY(XX,ZZ)
              ELSE
                ZZ=S18AEF(XX,IFAIL)
                ZZ=DLOG(ZZ)
                IF(IFAIL .NE. 0)THEN
                  PRINT*,'ERROR IN NAG:IFAIL=',IFAIL
                  STOP
                ENDIF
```

```
            ENDIF
            FC=FC+DLOG(Y)+DLOG(U(T(Num)))-Y*U(T(Num))
     *      *(RAIN(J)+RAIN(I)*(r**T(Num)))+(ZZ)
            Num=Num+1
            I=J
          ENDIF
30        CONTINUE
          FC=-FC-DLOG(Y)-DLOG(1.0D0-r)+(Y*(1.0D0-r)*RAIN(1))
          RETURN
C         End of function evaluation routine.
          END
C-------------------------------------------------------------------------
          SUBROUTINE POLY(XX,ZZ)
C         Polynomial approximation of I.(X) when the value of X
C         exceeds 87.9 where the nag routine can not compute.
          REAL*8 XX,ZZ,t
C
          t=3.75D0/XX
          ZZ=0.39894228D0+(0.01328592D0*t)+(0.00225319D0*(t**2))
     *    -(0.00157565D0*(t**3))+(0.00916281D0*(t**4))
     *    -(0.02057706D0*(t**5))+(0.02635537D0*(t**6))
     *    -(0.01647633D0*(t**7))+(0.00392377D0*(t**8))
          ZZ=DLOG(ZZ)-(0.5D0*DLOG(XX))+XX
          RETURN
          END
```